

(19)  **Europäisches Patentamt**
European Patent Office
Office européen des brevets



(11) **EP 1 229 134 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
 07.08.2002 Bulletin 2002/32

(51) Int Cl.7: **C12Q 1/68, C12N 15/10,
 G01N 33/50**

(21) Application number: **02250681.0**

(22) Date of filing: **31.01.2002**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
 MC NL PT SE TR**
 Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **31.01.2001 US 265805 P
 26.10.2001 US 339260 P**

(71) Applicants:
 • **Nucleonics, Inc**
Malvern, Pennsylvania 19355 (US)
 • **Message Pharmaceuticals, Inc.**
Malvern, PA 19355 (US)

(72) Inventors:
 • **Giordano, Tony**
Pheonixville, PA 19460 (US)
 • **Pachuk, Catherine**
Lansdale, PA 19446 (US)
 • **Satishchandran, Chandrasekhar**
Lansdale, PA 19446 (US)

(74) Representative: **Davies, Jonathan Mark**
Reddie & Grose
16 Theobalds Road
London WC1X 8PL (GB)

(54) **Use of post-transcriptional gene silencing for identifying nucleic acid sequences that modulate the function of a cell**

(57) Described herein are methods for identifying nucleic acid sequences that modulate the function of a cell, the expression of a gene in a cell, or the biological activity of a target polypeptide in a cell. The methods

involve the use of double stranded RNA expression libraries, double stranded RNA molecules, and post-transcriptional gene silencing techniques.

EP 1 229 134 A2

1

EP 1 229 134 A2

2

Description**Background of the Invention**

[0001] The invention relates to methods for identifying nucleic acid sequences that modulate the function of a cell, by the use of post-transcriptional gene silencing.

[0002] Double stranded RNA (dsRNA) has been shown to induce sequence-specific gene silencing in a number of different organisms. Gene silencing can occur through various mechanisms, one of which is post-transcriptional gene silencing (PTGS). In post-transcriptional gene silencing, transcription of the target locus is not affected, but the RNA half-life is decreased. The mechanisms by which PTGS occurs are not yet clear. Exogenous dsRNA has been shown to act as a potent inducer of PTGS in nematodes, trypanosomes, and insects. In addition, studies in *C. elegans* and *Drosophila* show that a few molecules of dsRNA per cell are sufficient to trigger a PTGS response. Furthermore, studies in mice have demonstrated that dsRNA can interfere with the expression of genes in mouse embryos.

[0003] There exists a need to identify molecules that selectively regulate the expression of genes in vertebrate cells without the associated toxicity of the interferon response. Such regulation should allow the down-regulation of expression from genes whose gene products are detrimental to the cells.

Summary of the Invention

[0004] In general the invention features high throughput methods of using PTGS to identify a nucleic acid sequence that modulates the function of a cell, gene expression of a target nucleic acid, or the biological activity of a target polypeptide. The method involves the use of specially constructed cDNA libraries derived from a cell, for example, a primary cell or a cell line that has an observable phenotype or biological activity, (e.g., an activity mediated by a target polypeptide or altered gene expression), that are transfected into cells to inhibit gene expression. This inhibition of gene expression alters the function of a cell, gene expression of a target nucleic acid, or the biological activity of a target polypeptide, and the nucleic acid sequence responsible for the modulation can be readily identified. The method may also utilize randomized nucleic acid sequences or a given sequence for which the function is not known. Although the use of PTGS as a validation strategy is known in the art, its use in screening techniques, as described herein, is novel.

[0005] Accordingly, in a first aspect, the invention features a method for identifying a nucleic acid sequence that modulates the function of a cell. The method involves: (a) transforming a population of cells with a double stranded RNA expression library, where the library is derived from the cells, where at least two cells of the population of cells are each transformed with a different

nucleic acid from the double stranded RNA expression library, and where the nucleic acid is capable of forming double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed in the cell; and (c) assaying for a modulation in the function of the cell, wherein a modulation identifies a nucleic acid sequence that modulates the function of a cell.

[0006] In a desirable embodiment of the first aspect of the invention, assaying for a modulation in the function of a cell comprises measuring cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, or the ability of a cell to support viral replication.

[0007] In a second aspect, the invention features a method for identifying a nucleic acid sequence that modulates expression of a target nucleic acid in a cell. The method involves: (a) transforming a population of cells with a double stranded RNA expression library, where the library is derived from the cells, where at least two cells of the population of cells are each transformed with a different nucleic acid from the double stranded RNA expression library, and where the nucleic acid is capable of forming double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed in the cell; and (c) assaying for a modulation in the expression of a gene in the cell, where a modulation identifies a nucleic acid sequence that modulates expression of a target nucleic acid in a cell.

[0008] In a desirable embodiment of the second aspect of the invention, the target nucleic acid is assayed using DNA array technology.

[0009] In a third aspect, the invention features a method for identifying a nucleic acid sequence that modulates the biological activity of a target polypeptide in a cell. The method involves: (a) transforming a population of cells with a double stranded RNA expression library, where the library is derived from the cells, where at least two cells of the population of cells are each transformed with a different nucleic acid from the double stranded RNA expression library, and where the nucleic acid is capable of forming double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed in the cell; and (c) assaying for a modulation in the biological activity of a target polypeptide in the cell, wherein a modulation identifies a nucleic acid sequence that modulates the biological activity of a target polypeptide.

[0010] In one embodiment of any of the above aspects of the invention, in transforming step (a), the nucleic acid is stably integrated into a chromosome of the cell. Integration of the nucleic acid may be random or site-specific. Desirably integration is mediated by recombination or retroviral insertion. In addition, desirably a single copy of the nucleic acid is integrated into the chromosome. In another embodiment of any of the above aspects of the invention, in step (a) at least 50, more desirably 100; 500; 1000; 10,000; or 50,000 cells of the population of cells are each transformed with a

3

EP 1 229 134 A2

4

different nucleic acid from the double stranded RNA expression library. In other embodiments, the population of cells is transformed with at least 5%, more desirably at least 25%, 50%, 75%, or 90%, and most desirably at least 95% of the double stranded RNA expression library. In yet another embodiment, the method further involves: (d) identifying the nucleic acid sequence by amplifying and cloning the sequence. Desirably amplification of the sequence involves the use of the polymerase chain reaction (PCR).

[0011] In other embodiments of any of the above aspects of the invention, the double stranded RNA expression library contains cDNAs or randomized nucleic acids. The double stranded RNA expression library may be a nuclear double stranded RNA expression library, in which case the double stranded nucleic acid is made in the nucleus. Alternatively, the double stranded RNA expression library may be a cytoplasmic double stranded RNA expression library, in which case the double stranded nucleic acid is made in the cytoplasm. In addition, the nucleic acid from the double stranded RNA expression library may be made *in vitro* or *in vivo*. In addition, the identified nucleic acid sequence may be located in the cytoplasm of the cell.

In still another embodiment of any of the above aspects of the invention, the nucleic acid is contained in a vector, for example a double stranded RNA expression vector. The vector may then be transformed such that it is stably integrated into a chromosome of the cell, or it may function as an episomal (non-integrated) expression vector within the cell. In one embodiment, a vector that is integrated into a chromosome of the cell contains a promoter operably linked to a nucleic acid encoding a hairpin or double stranded RNA. In another embodiment, the vector does not contain a promoter operably linked to a nucleic acid encoding a double stranded RNA. In this latter embodiment, the vector integrates into a chromosome of a cell such that an endogenous promoter is operably linked to a nucleic acid from the vector that encodes a double stranded RNA. Desirably, the double stranded RNA expression vector comprises at least one RNA polymerase II promoter, for example, a human CMV-immediate early promoter (HCMV-IE) or a simian CMV (SCMV) promoter, at least one RNA polymerase I promoter, or at least one RNA polymerase III promoter. The promoter may also be a T7 promoter, in which case, the cell further comprises T7 polymerase. Alternatively, the promoter may be an SP6 promoter, in which case, the cell further comprises SP6 polymerase. The promoter may also be one convergent T7 promoter and one convergent SP6 promoter. A cell may be made to contain T7 or SP6 polymerase by transforming the cell with a T7 polymerase or an SP6 polymerase expression plasmid, respectively. In some embodiments, a T7 promoter or a RNA polymerase III promoter is operably linked to a nucleic acid that encodes a small double stranded RNA (e.g., a double stranded RNA that is less than 200, 150, 100, 75, 50, or 25 nucleotides in length). In other

embodiments, the promoter is a mitochondrial promoter that allows cytoplasmic transcription of the nucleic acid in the vector (see, for example, the mitochondrial promoters described in WO 00/63364, filed April 19, 2000).

Alternatively, the promoter is an inducible promoter, such as a *lac* (Cronin *et al.* *Genes & Development* 15: 1506-1517, 2001), *ara* (Khlebnikov *et al.*, *J. Bacteriol.* 2000 Dec;182(24):7029-34), ecdysone (Rheogene, www.rheogene.com), RU48 (mefepiristone) (corticosteroid antagonist) (Wang XJ, Liefer KM, Tsai S, O'Malley BW, Roop DR, *Proc Natl Acad Sci U S A.* 1999 Jul 20; 96(15):8483-8), or *tet* promoter (Rendal *et al.*, *Hum Gene Ther.* 2002 Jan;13(2):335-42. and Larnarina *et al.*, *Hum Gene Ther.* 2002 Jan;13(2):199-210) or a promoter disclosed in WO 00/63364, filed April 19, 2000. In desirable embodiments, the inducible promoter is not induced until all the episomal vectors are eliminated from the cell. The vector may also comprise a selectable marker. In addition, these vectors may be used in combination with methods that inhibit or prevent an interferon response or double stranded RNA stress response, as described herein.

[0012] Desirably in a vector for use in any of the above aspects of the invention, the sense strand and the antisense strand of the nucleic acid sequence are transcribed from the same nucleic acid sequence using two convergent promoters. In another desirable embodiment, in a vector for use in any of the above aspects of the invention, the nucleic acid sequence comprises an inverted repeat, such that upon transcription, the nucleic acid forms a double stranded RNA.

[0013] In still other embodiments of any of the above aspects of the invention, the cell and the vector each further comprise a *loxP* site and site-specific integration of the nucleic acid into a chromosome of the cell occurs through recombination between the *loxP* sites. In addition, step (b) of any of the above aspects of the invention further involves rescuing the nucleic acid through Cre-mediated double recombination.

[0014] In still further embodiments of any of the above aspects of the invention, the identified nucleic acid sequence is located in the nucleus of the cell. Alternatively, the identified nucleic acid sequence may be located in the cytoplasm of the cell.

[0015] In yet another embodiment of any of the above aspects of the invention, the nucleic acid from the double stranded RNA expression library is at least 100, 500, 600, or 1000 nucleotides in length. In other embodiments of any of the above aspects of the invention, the nucleic acid from the double stranded RNA expression library is at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the nucleic acid from the double stranded RNA expression library is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucle-

5

EP 1 229 134 A2

6

otides in the nucleic acid from the double stranded RNA expression library is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the nucleic acid contains less than 50,000; 10,000; 5,000; or 2,000 nucleotides. In addition, the nucleic acid from the double stranded RNA expression library may contain a sequence that is less than a full length RNA sequence.

[0016] In still further embodiments of any of the above aspects of the invention, the cell is a plant cell or an animal cell. Desirably the animal cell is a vertebrate or mammalian cell, for example, a human cell. The cell may be *ex vivo* or *in vivo*. The cell may be a gamete or a somatic cell, for example, a cancer cell, a stem cell, a cell of the immune system, a neuronal cell, a muscle cell, or an adipocyte.

[0017] Transformation/transfection of the cell may occur through a variety of means including, but not limited to, lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, viral or retroviral delivery, electroporation, or biolistic transformation. The RNA or RNA expression vector (DNA) may be naked RNA or DNA or local anesthetic complexed RNA or DNA (Pachuk *et al.*, Biochim. Biophys. Acta 1468:20-30, 2000). In another embodiment, the cell is not a *C. elegans* cell. Desirably the vertebrate or mammalian cell has been cultured for only a small number of passages (*e.g.*, less than 30 passages of a cell line that has been directly obtained from American Type Culture Collection), or are primary cells. Desirably, the vertebrate or mammalian cell is transformed with nucleic acids that are not complexed with cationic lipids.

[0018] In yet another embodiment of any of the above aspects of the invention, the cell is derived from a parent cell, and is generated by: (a) transforming a population of parent cells with a bicistronic plasmid expressing a selectable marker and a reporter gene, and comprising a *loxP* site; (b) selecting for a cell in which the plasmid is stably integrated; and (c) selecting for a cell in which one copy of the plasmid is stably integrated in a transcriptionally active locus. Desirably the selectable marker is G418 and the reporter gene is green fluorescent protein (GFP).

[0019] In still another embodiment of the above aspects of the invention, generation of the double stranded expression library comprises: (a) isolating RNA from a cell; (b) synthesizing cDNAs from the RNA of step (a); and (c) cloning each cDNA into a vector. Desirably cDNA synthesis is optimized and/or size selected for the generation and/or selection of cDNAs that are at least 100, 500, 600, or 1000 nucleotides in length. In other embodiments, the cDNAs are least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the cDNAs is

between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the cDNAs is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the cDNAs contain less than 50,000; 10,000; 5,000; or 2,000 nucleotides. In addition, the cDNA may encode an RNA fragment that is less than full length. Desirably the vector comprises two convergent T7 promoters, two convergent SP6 promoters, or one convergent T7 promoter and one convergent SP6 promoter, a selectable marker, and/or a *loxP* site.

[0020] In an additional embodiment of any of the above aspects of the invention, the method is carried out under conditions that inhibit or prevent an interferon response or double stranded RNA stress response.

[0021] In a fourth aspect, the invention features a method for identifying a nucleic acid sequence that modulates the function of a cell, involving: (a) transforming a population of cells with a double stranded RNA that is derived from the cells; (b) optionally selecting for a cell in which the nucleic acid is expressed; and (c) assaying for a modulation in the function of the cell, wherein the modulation identifies a nucleic acid sequence that modulates the function of a cell, wherein the method is desirably carried out under conditions that inhibit or prevent an interferon response or double stranded RNA stress response.

[0022] In a desirable embodiment of the fourth aspect of the invention, assaying for a modulation in the function of a cell comprises measuring cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, or the ability of a cell to support viral replication.

[0023] In a fifth aspect, the invention features a method for identifying a nucleic acid sequence that modulates expression of a target nucleic acid in a cell, involving: (a) transforming a population of cells with a double stranded RNA that is derived from the cells; (b) optionally selecting for a cell in which the nucleic acid is expressed; and (c) assaying for a modulation in the expression of the gene in the cell, wherein the modulation identifies a nucleic acid sequence that modulates expression of a target nucleic acid in a cell, wherein the method is desirably carried out under conditions that inhibit or prevent an interferon response or double stranded RNA stress response.

[0024] In a desirable embodiment of the fifth aspect of the invention, the target nucleic acid is assayed using DNA array technology.

[0025] In a sixth aspect, the invention features a method for identifying a nucleic acid sequence that modulates the biological activity of a target polypeptide in a

cell, involving: (a) transforming a population of cells with a double stranded RNA that is derived from the cells; (b) optionally selecting for a cell in which the nucleic acid is expressed in the cell; and (c) assaying for a modulation in the biological activity of a target polypeptide in the cell, wherein the modulation identifies a nucleic acid sequence that modulates the biological activity of a target polypeptide in a cell, wherein the method is desirably carried out under conditions that inhibit or prevent an interferon response or double stranded RNA stress response.

[0026] In a seventh aspect, the invention features a method for identifying a nucleic acid sequence that modulates the function of a cell, involving: (a) transforming a population of cells with a double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed; and (c) assaying for a modulation in the function of the cell. Desirably, the modulation identifies a nucleic acid sequence that modulates the function of a cell, wherein the method is desirably carried out under conditions that or prevent an interferon response or double stranded RNA stress response.

[0027] In a desirable embodiment of the seventh aspect of the invention, assaying for a modulation in the function of a cell comprises measuring cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, or the ability of a cell to support viral replication.

[0028] In a eighth aspect, the invention features a method for identifying a nucleic acid sequence that modulates expression of a target nucleic acid in a cell, involving: (a) transforming a population of cells with a double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed; and (c) assaying for a modulation in the expression of the gene in the cell, wherein the modulation identifies a nucleic acid sequence that modulates expression of a target nucleic acid in a cell. Desirably, the method is carried out under conditions that inhibit or prevent an interferon response or double stranded RNA stress response.

[0029] In a desirable embodiment of the eighth aspect of the invention, the target nucleic acid is assayed using DNA array technology.

[0030] In a ninth aspect, the invention features a method for identifying a nucleic acid sequence that modulates the biological activity of a target polypeptide in a cell, involving: (a) transforming a population of cells with a double stranded RNA; (b) optionally selecting for a cell in which the nucleic acid is expressed in the cell; and (c) assaying for a modulation in the biological activity of a target polypeptide in the cell, wherein the modulation identifies a nucleic acid sequence that modulates the biological activity of a target polypeptide in a cell. Desirably, the method is carried out under conditions that inhibit or prevent an interferon response double stranded RNA stress response.

[0031] In one embodiment of any of the above as-

pects of the invention, in step (a) at least 2, more desirably 50; 100; 500; 1000; 10,000; or 50,000 cells of the population of cells are each transformed with a different double stranded RNA from a double stranded RNA expression library. Desirably, at most one double stranded RNA is inserted into each cell. In other embodiments, the population of cells is transformed with at least 5%, more desirably at least 25%, 50%, 75%, or 90%, and most desirably, at least 95% of the double stranded RNA expression library. In still another embodiment of any of the fourth, fifth, or sixth aspects of the invention, the method further involves: (d) identifying the nucleic acid sequence by amplifying and cloning the sequence. Desirably amplification of the sequence involves the use of the polymerase chain reaction (PCR).

[0032] In a tenth aspect, the invention features a cell or a population of cells that expresses a double stranded RNA that (i) modulates a function of the cell, (ii) modulates the expression of a target nucleic acid (e.g., an endogenous or pathogen gene) in the cell, and/or (iii) modulates the biological activity of a target protein (e.g., an endogenous or pathogen protein) in the cell. Desirably, the cell contains only one molecular species of double stranded RNA or only one copy of a double stranded RNA expression vector (e.g., a stably integrated vector). Desirably, the cell or population of cells is produced using one or more methods of the invention. In other embodiments, the double stranded RNA is expressed under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response.

[0033] In other embodiments of any of the fourth, fifth, sixth, seventh, eighth, ninth, or tenth aspects of the invention, the double stranded RNA is derived from cDNAs or randomized nucleic acids. In addition, the double stranded RNA may be a cytoplasmic double stranded RNA, in which case the double stranded nucleic acid is made in the cytoplasm. The double stranded RNA may be made *in vitro* or *in vivo*. In addition, the identified nucleic acid sequence may be located in the cytoplasm of the cell.

[0034] In still another embodiment of any of the fourth, fifth, sixth, seventh, eighth, ninth, or tenth aspects of the invention, the nucleic acid is contained in a vector, for example, a double stranded RNA expression vector that is capable of forming a double stranded RNA. Desirably the double stranded RNA expression vector comprises at least one promoter. The promoter may be a T7 promoter, in which case, the cell further comprises T7 polymerase. Alternatively, the promoter may be an SP6 promoter, in which case, the cell further comprises SP6 polymerase. The promoter may also be one convergent T7 promoter and one convergent SP6 promoter. A cell may be made to contain T7 or SP6 polymerase by transforming the cell with a T7 polymerase or an SP6 polymerase expression plasmid, respectively. The vector may also comprise a selectable marker, for example hygromycin.

[0035] Desirably in a vector for use in any of the fourth,

fifth, sixth, seventh, eighth, ninth, or tenth aspects of the invention, the sense strand and the antisense strand of the nucleic acid sequence are transcribed from the same nucleic acid sequence using two convergent promoters. In another desirable embodiment, in a vector for use in any of the above aspects of the invention, the nucleic acid sequence comprises an inverted repeat, such that upon transcription, the nucleic acid forms a double stranded RNA.

[0036] In yet another embodiment of any of the fourth, fifth, sixth, seventh, eighth, ninth, or tenth aspects of the invention, the double stranded RNA is at least 100, 500, 600, or 1000 nucleotides in length. In other embodiments of any of the fourth, fifth, or sixth aspects of the invention, the double stranded RNA is at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the double stranded RNA is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the double stranded RNA is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the double stranded RNA contains less than 50,000; 10,000; 5,000; or 2,000 nucleotides. In addition, the double stranded RNA may contain a sequence that is less than a full length RNA sequence.

[0037] In still further embodiments of any of the fourth, fifth, sixth, seventh, eighth, ninth, or tenth aspects of the invention, the cell is a plant cell or an animal cell. Desirably the animal cell is a vertebrate or mammalian cell, for example, a human cell. The cell may be *ex vivo* or *in vivo*. The cell may be a gamete or a somatic cell, for example, a cancer cell, a stem cell, a cell of the immune system, a neuronal cell, a muscle cell, or an adipocyte.

[0038] In other embodiments of any of the first, second, third, seventh, eighth, ninth, or tenth aspects of the invention, the double stranded RNA is derived from a cell or a population of cells and used to transform another cell population of either the same cell type or a different cell type. In desirable embodiments, the transformed cell population contains cells of a cell type that is related to the cell type of the cells from which the double stranded RNA was derived (*e.g.*, the transformation of cells of one neuronal cell type with the double stranded RNA derived from cells of another neuronal cell type). In yet other embodiments of any of these aspects, the double stranded RNA contains one or more contiguous or non-contiguous positions that are randomized (*e.g.*, by chemical or enzymatic synthesis using a mixture of nucleotides that may be added at the randomized position). In still other embodiments, the double stranded RNA is a randomized nucleic acid in which segments of ribonucleotides and/or deoxyribonucleotides are ligated

to form the double stranded RNA.

[0039] In other embodiments of any of various aspects of the invention, the double stranded RNA specifically hybridizes to a target nucleic acid but does not substantially hybridize to non-target molecules, which include other nucleic acids in the cell or biological sample having a sequence that is less than 99, 95, 90, 80, or 70% identical or complementary to that of the target nucleic acid. Desirably, the amount of the these non-target molecules hybridized to, or associated with, the double stranded RNA, as measured using standard assays, is 2-fold, desirably 5-fold, more desirably 10-fold, and most desirably 50-fold lower than the amount of the target nucleic acid hybridized to, or associated with, the double stranded RNA. In other embodiments, the amount of a target nucleic acid hybridized to, or associated with, the double stranded RNA, as measured using standard assays, is 2-fold, desirably 5-fold, more desirably 10-fold, and most desirably 50-fold greater than the amount of a control nucleic acid hybridized to, or associated with, the double stranded RNA. Desirably, the double stranded RNA only hybridizes to one target nucleic acid from a cell under denaturing, high stringency hybridization conditions. In certain embodiments, the double stranded RNA is substantially homologous (*e.g.*, at least 80, 90, 95, 98, or 100% homologous) to only one target nucleic acid from a cell. In other embodiments, the double stranded RNA is homologous to multiple RNAs, such as RNAs from the same gene family. In yet other embodiments, the double stranded RNA is homologous to distinctly different mRNA sequences from genes that are similarly regulated (*e.g.*, developmental, chromatin remodeling, or stress response induced). In other embodiments, the double stranded RNA is homologous to a large number of RNA molecules, such as a double stranded RNA designed to induce a stress response or apoptosis. In other embodiments, the percent decrease in the expression of a target nucleic acid is at least 2, 5, 10, 20, or 50 fold greater than the percent decrease in the expression of a non-target or control nucleic acid. Desirably, the double stranded RNA inhibits the expression of a target nucleic acid but has negligible, if any, effect on the expression of other nucleic acids in the cell. Examples of control nucleic acids include nucleic acids with a random sequence or nucleic acids known to have little, if any, affinity for the double stranded RNA.

[0040] In other embodiments of any of various aspects of the invention, at most one molecular species of double stranded RNA is inserted into each cell. In other embodiments, at most one vector is stably integrated into the genome of each cell. In various embodiments, the double stranded RNA is active in the nucleus of the transformed cell and/or is active in the cytoplasm of the transformed cell. In various embodiments, at least 1, 10, 20, 50, 100, 500, or 1000 cells or all of the cells in the population are selected as cells that contain or express a double stranded RNA. In some embodiments, at least

1, 10, 20, 50, 100, 500, or 1000 cells or all of the cells in the population are assayed for a modulation in the function of the cell, a modulation in the expression of a target nucleic acid (e.g., an endogenous or pathogen gene) in the cell, and/or a modulation in the biological activity of a target protein (e.g., an endogenous or pathogen protein) in the cell.

[0041] In other embodiments, the double stranded RNA or double stranded RNA expression vector is complexed with one or more cationic lipids or cationic amphiphiles, such as the compositions disclosed in US 4,897,355 (Eppstein *et al.*, filed October 29, 1987), US 5,264,618 (Feigner *et al.*, filed April 16, 1991) or US 5,459,127 (Feigner *et al.*, filed September 16, 1993). In other embodiments, the double stranded RNA or double stranded RNA expression vector is complexed with a liposomes/liposomic composition that includes a cationic lipid and optionally includes another component such as a neutral lipid (see, for example, US 5,279,833 (Rose), US 5,283,185 (Eband), and US 5,932,241). In yet other embodiments, the double stranded RNA or double stranded RNA expression vector is complexed with any other composition that is devised by one of ordinary skill in the fields of pharmaceuticals and molecular biology.

[0042] Desirably, the double stranded RNA specifically hybridizes to a target nucleic acid but does not substantially hybridize to non-target molecules, which include other nucleic acids in the cell or biological sample having a sequence that is less than 99, 95, 90, 80, or 70 % identical to or complementary to that of the target nucleic acid. In other embodiments, the percent decrease in the expression of a target nucleic acid is at least 2, 5, 10, 20, or 50 fold greater than the percent decrease in the expression of a non-target or control nucleic acid. Desirably, the double stranded RNA inhibits the expression of the target nucleic acid but has negligible, if any, effect on the expression of other nucleic acids in the cell.

[0043] Transformation/transfection of the cell may occur through a variety of means including, but not limited to, lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, viral or retroviral delivery, electroporation, or biolistic transformation. The RNA or RNA expression vector (DNA) may be naked RNA or DNA or local anesthetic complexed RNA or DNA (Pachuk *et al.*, *supra*). In yet another embodiment, the cell is not a *C. elegans* cell. Desirably the vertebrate or mammalian cell has been cultured for only a small number of passages (e.g., less than 30 passages of a cell line that has been directly obtained from American Type Culture Collection), or are primary cells. In addition, desirably the vertebrate or mammalian cell is transformed with double stranded RNA that is not complexed with cationic lipids.

[0044] The transcription systems described herein provide advantages to other double stranded expression systems. Following transformation of the double

stranded RNA library, cells contain hundreds to thousands of double stranded RNA expression cassettes, with concomitant expression of that many expression cassettes. In the double stranded RNA expression system of the present invention, double stranded RNA (dsRNA) expression cassettes contained within the expression vector integrate into the chromosome of the transfected cell. Desirably, every transformed cell integrates one of the double stranded expression cassettes.

Through expansion of the transformed cell, episomal (non-integrated) expression vectors are diluted out of the cell over time. Desirably no transcription occurs until the episomal expression vectors are diluted out of the cell, such that not more than 5 episomal vectors remain in the cell. Most desirably, no transcription occurs until all of the episomal vectors have been diluted out of the cell, and only the integrated expression cassette remains. The time it takes for all episomal vectors to be removed from the cell is proportional to the replication rate of the transformed cell, and is generally on the order of two to several weeks of cell culture and growth. The numbers of copies of a dsRNA molecule in a transformed cell can be determined using, for example, standard PCR techniques, and thereby, the number of episomal vectors in a given cell can be monitored.

[0045] Once a stable integrant containing five or fewer, and desirably no episomal expression vectors, transcription is induced, allowing dsRNA to be expressed in the cells. This method ensures that, if desired, only one species or not more than about five species of dsRNA is expressed per cell, as opposed to other methods that express hundreds to thousands of double stranded species.

[0046] Another problem that can occur in other double stranded expression systems or dsRNA delivery systems is that some dsRNA sequences, possibly in certain cell types and through certain delivery methods, may result in an interferon response (Jaramillo *et al.*, *Cancer Invest.* 13:327-338, 1995). During the induction of post-transcriptional gene silencing events, induction of an interferon response is not desired, as this could lead to cell death and possibly to the prevention of gene silencing. An additional advantage of the present invention is that the dsRNA delivery methods described herein are performed such that an interferon response is inhibited or prevented.

[0047] One of the components of an interferon response is the induction of the interferon-induced protein kinase PKR (Jaramillo *et al.*, *supra*). Suppression of the interferon response and/or the PKR response, using techniques described herein, is desired in the cells targeted for a PTGS event in those instances where an interferon response would otherwise be induced. Methods for suppressing an interferon response or dsRNA stress response can be used in combination with any of the methods for identifying a nucleic acid sequence that modulates the function of a cell, gene expression in a cell, or the biological activity of a target polypeptide.

13

EP 1 229 134 A2

14

[0048] The methods of the present invention provide a means for high throughput identification of nucleic acid sequences involved in modulating the function of a cell, the expression of a target nucleic acid in a cell, or the biological activity of a target polypeptide in a cell. By transforming a population of cells with a double stranded RNA expression library, the effects of many PTGS events on cell function, expression of a target nucleic acid in a cell, or the biological activity of a target polypeptide in a cell can be evaluated simultaneously, thereby allowing for rapid identification of the nucleic acid sequence involved in a cell function, target nucleic acid expression, or biological activity of a target polypeptide of interest.

[0049] By "nucleic acid," "nucleic acid sequence," "double stranded RNA nucleic acid sequence," or "double stranded RNA nucleic acid" is meant a nucleic acid or a portion thereof that is free of the genes that, in the naturally-occurring genome of the organism from which the nucleic acid sequence of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA, with or without 5' or 3' flanking sequences that is incorporated into a vector, for example, a double stranded RNA expression vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences.

[0050] By "double stranded RNA" is meant a nucleic acid containing a region of two or more nucleotides that are in a double stranded conformation. In various embodiments, the double stranded RNA consists entirely of ribonucleotides or consists of a mixture of ribonucleotides and deoxynucleotides, such as the RNA/DNA hybrids disclosed, for example, by WO 00/63364, filed April 19, 2000 or U.S.S.N. 60/130,377, filed April 21, 1999. The double stranded RNA may be a single molecule with a region of self-complementarity such that nucleotides in one segment of the molecule base pair with nucleotides in another segment of the molecule. In various embodiments, a double stranded RNA that consists of a single molecule consists entirely of ribonucleotides or includes a region of ribonucleotides that is complementary to a region of deoxyribonucleotides. Alternatively, the double stranded RNA may include two different strands that have a region of complementarity to each other. In various embodiments, both strands consist entirely of ribonucleotides, one strand consists entirely of ribonucleotides and one strand consists entirely of deoxyribonucleotides, or one or both strands contain a mixture of ribonucleotides and deoxyribonucleotides. Desirably, the regions of complementarity are at least 70, 80, 90, 95, 98, or 100% complementary. Desirably, the region of the double stranded RNA that is present in a double stranded conformation includes at least 5, 10, 20, 30, 50, 75, 100, 200, 500, 1000, 2000 or 5000 nucleotides or includes all of the nucleotides in a cDNA being

represented in the double stranded RNA. In some embodiments, the double stranded RNA does not contain any single stranded regions, such as single stranded ends, or the double stranded RNA is a hairpin. Desirable RNA/DNA hybrids include a DNA strand or region that is an antisense strand or region (e.g., has at least 70, 80, 90, 95, 98, or 100% complementary to a target nucleic acid) and an RNA strand or region that is a sense strand or region (e.g., has at least 70, 80, 90, 95, 98, or 100% identity to a target nucleic acid). In various embodiments, the RNA/DNA hybrid is made *in vitro* using enzymatic or chemical synthetic methods such as those described herein or those described in WO 00/63364, filed April 19, 2000 or U.S.S.N. 60/130,377, filed April 21, 1999. In other embodiments, a DNA strand synthesized *in vitro* is complexed with an RNA strand made *in vivo* or *in vitro* before, after, or concurrent with the transformation of the DNA strand into the cell. In yet other embodiments, the double stranded RNA is a single circular nucleic acid containing a sense and an antisense region, or the double stranded RNA includes a circular nucleic acid and either a second circular nucleic acid or a linear nucleic acid (see, for example, WO 00/63364, filed April 19, 2000 or U.S.S.N. 60/130,377, filed April 21, 1999.) Exemplary circular nucleic acids include lariat structures in which the free 5' phosphoryl group of a nucleotide becomes linked to the 2' hydroxyl group of another nucleotide in a loop back fashion.

[0051] In other embodiments, the double stranded RNA includes one or more modified nucleotides in which the 2' position in the sugar contains a halogen (such as fluorine group) or contains an alkoxy group (such as a methoxy group) which increases the half-life of the double stranded RNA *in vitro* or *in vivo* compared to the corresponding double stranded RNA in which the corresponding 2' position contains a hydrogen or an hydroxyl group. In yet other embodiments, the double stranded RNA includes one or more linkages between adjacent nucleotides other than a naturally-occurring phosphodiester linkage. Examples of such linkages include phosphoramidate, phosphorothioate, and phosphorodithioate linkages. In other embodiments, the double stranded RNA contains one or two capped strands, as disclosed, for example, by WO 00/63364, filed April 19, 2000 or U.S.S.N. 60/130,377, filed April 21, 1999. In other embodiments, the double stranded RNA contains coding sequence or non-coding sequence, for example, a regulatory sequence (e.g., a transcription factor binding site, a promoter, or a 5' or 3' untranslated region (UTR) of an mRNA). Additionally, the double stranded RNA can be any of the at least partially double-stranded RNA molecules disclosed in WO 00/63364, filed April 19, 2000 (see, for example, pages 8-22). Any of the double stranded RNAs may be expressed *in vitro* or *in vivo* using the methods described herein or standard methods, such as those described in WO 00/63364, filed April 19, 2000 (see, for example, pages 16-22).

[0052] By "double stranded RNA expression library"

15

EP 1 229 134 A2

16

or "dsRNA expression library" is meant a collection of nucleic acid expression vectors containing nucleic acid sequences, for example, cDNA sequences or randomized nucleic acid sequences that are capable of forming a double stranded RNA (dsRNA) upon expression of the nucleic acid sequence. Desirably the double stranded RNA expression library contains at least 10,000 unique nucleic acid sequences, more desirably at least 50,000; 100,000; or 500,000 unique nucleic acid sequences, and most desirably, at least 1,000,000 unique nucleic acid sequences. By a "unique nucleic acid sequence" is meant that a nucleic acid sequence of a double stranded RNA expression library has desirably less than 50%, more desirably less than 25% or 20%, and most desirably less than 10% nucleic acid identity to another nucleic acid sequence of a double stranded RNA expression library when the full length sequence are compared. Sequence identity is typically measured using sequence analysis software with the default parameters specified therein (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). This software program matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

[0053] The preparation of cDNAs for the generation of double stranded RNA expression libraries is described herein. A randomized nucleic acid library may also be generated as described in detail below. The double stranded RNA expression library may contain nucleic acid sequences that are transcribed in the nucleus or that are transcribed in the cytoplasm of the cell. A double stranded RNA expression library may be generated using techniques described herein.

[0054] By "target nucleic acid" is meant a nucleic acid sequence whose expression is modulated as a result of post-transcriptional gene silencing. As used herein, the target nucleic acid may be in the cell in which the PTGS event occurs or it may be in a neighboring cell, or in a cell contacted with media or other extracellular fluid in which the cell that has undergone the PTGS event is contained. Exemplary target nucleic acids include nucleic acids associated with cancer or abnormal cell growth, such as oncogenes, and nucleic acids associated with an autosomal dominant or recessive disorder. Desirably, the double stranded RNA inhibits the expression of an allele of a nucleic acid that has a mutation associated with a dominant disorder and does not substantially inhibit the other allele of the nucleic acid (e.g., an allele without a mutation associated with the disorder). Other exemplary target nucleic acids include host cellular nucleic acids or pathogen nucleic acids required for the infection or propagation of a pathogen, such as a virus, bacteria, yeast, protozoa, or parasite.

[0055] By "target polypeptide" is meant a polypeptide whose biological activity is modulated as a result of post-transcriptional gene silencing. As used herein, the target

polypeptide may be in the cell in which the PTGS event occurs or it may be in a neighboring cell, or in a cell contacted with media or other extracellular fluid in which the cell that has undergone the PTGS event is contained.

[0056] As used herein, by "randomized nucleic acids" is meant nucleic acids, for example, those that are at least 100, 500, 600, or 1000 nucleotides in length, constructed from RNA isolated from a particular cell type. In other embodiments, the nucleic acids are at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the nucleic acids is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the nucleic acids is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the nucleic acids contain less than 50,000; 10,000; 5,000; or 2,000 nucleotides. A randomized nucleic acid library may be constructed in a number of ways. For example, it may be constructed from existing cDNA libraries. In one example, the cDNA libraries are shuffled using the "Gene Shuffling" technology of Maxygen Corp. The cDNA sequences are amplified using inefficient PCR either by restricting elongation time or through the use of manganese. A library of recombinants is created, and the library is finally amplified by PCR and cloned into vectors. In a second method, existing cDNA libraries are digested with an endonuclease to generate fragments of 10 to 300 base pairs. Alternatively, the cDNA libraries are digested to generate shorter fragments of, for example, 5 to 50 base pairs, 5 to 40 base pairs, 5 to 20 base pairs, 5 to 10 base pairs, or 10 to 20 base pairs, inclusive. If the fragments are to contain 5' OH and 3' PO₄ groups, they are dephosphorylated using alkaline phosphatase and phosphorylated using polynucleotide kinase. These dsDNA fragments are then ligated to form larger molecules, and are size selected. In a third example, randomized nucleic acid libraries are created by using random priming of cDNA libraries (using random hexamers and Klenow) to generate short fragments of 20 to 100 nucleotides. Alternatively, shorter fragments are generated that contain, for example, 5 to 50 nucleotides, 5 to 40 nucleotides, 5 to 20 nucleotides, 5 to 10 nucleotides, or 10 to 20 nucleotides, inclusive. These fragments are then ligated randomly to give a desired sized larger fragment.

[0057] Alternatively, a randomized nucleic acid library can be generated from random sequences of oligonucleotides. For example, DNA or RNA oligonucleotides may be prepared chemically. Random DNA sequences may also be prepared enzymatically using terminal transferase in the presence of all dNTPs. Random RNA molecules may be prepared using NDPs and NDP phos-

17

EP 1 229 134 A2

18

phorylase. The random sequences may be 10 to 300 bases in length. Alternatively, shorter random sequences are used that contain, for example, 5 to 50 bases, 5 to 40 bases, 5 to 20 bases, 5 to 10 bases, or 10 to 20 bases, inclusive. The sequences are ligated to form the desired larger sequence using RNA ligase. Alternatively these sequences may be ligated chemically. The oligonucleotides are phosphorylated at the 5' position using polynucleotide kinase or by chemical methods, prior to ligation enzymatically. Chemical ligations can utilize a 5' PO₄ and a 3' OH group or a 5' OH and a 3' PO₄ group.

[0058] Alternatively, a randomized nucleic acid library can be generated by converting the random DNA sequences into dsDNA sequences using DNA polymerase (Klenow), dNTP and random heteromeric primers, and the RNA sequences are converted into dsDNA sequences by reverse transcriptase and Klenow. After converting into DNA (ss or ds) the sequences are then amplified by PCR. The dsDNA fragments can also be ligated to give larger fragments of a desired size.

[0059] The randomized nucleic acids may be cloned into a vector, for example, an expression vector, as a double stranded RNA transcription cassette. The sequence of the nucleic acid may not be known at the time the vector is generated. The randomized nucleic acid may contain coding sequence or non-coding sequence, for example, a regulatory sequence (e.g., a transcription factor binding site, a promoter, or a 5' or 3' untranslated region (UTR) of an mRNA).

[0060] By "Cre-mediated double recombination" is meant two nucleic acid recombination events involving *loxP* sites that are mediated by Cre recombinase. A Cre-mediated double recombination event can occur, for example, as illustrated in Fig. 1.

[0061] By "function of a cell" is meant any cell activity that can be measured or assessed. Examples of cell function include, but are not limited to, cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, and the ability of a cell to support viral replication. The function of a cell may also be to affect the function, gene expression, or the polypeptide biological activity of another cell, for example, a neighboring cell, a cell that is contacted with the cell, or a cell that is contacted with media or other extracellular fluid that the cell is contained in.

[0062] By "apoptosis" is meant a cell death pathway wherein a dying cell displays a set of well-characterized biochemical hallmarks that include cytolemmal membrane blebbing, cell soma shrinkage, chromatin condensation, nuclear disintegration, and DNA laddering. There are many well-known assays for determining the apoptotic state of a cell, including, and not limited to: reduction of MTT tetrazolium dye, TUNEL staining, Annexin V staining, propidium iodide staining, DNA laddering, PARP cleavage, caspase activation, and assessment of cellular and nuclear morphology. Any of these or other known assays may be used in the methods of

the invention to determine whether a cell is undergoing apoptosis.

[0063] By "polypeptide biological activity" is meant the ability of a target polypeptide to modulate cell function.

5 The level of polypeptide biological activity may be directly measured using standard assays known in the art. For example, the relative level of polypeptide biological activity may be assessed by measuring the level of the mRNA that encodes the target polypeptide (e.g., by reverse transcription-polymerase chain reaction (RT-PCR) amplification or Northern blot analysis); the level of target polypeptide (e.g., by ELISA or Western blot analysis); the activity of a reporter gene under the transcriptional regulation of a target polypeptide transcriptional regulatory region (e.g., by reporter gene assay, as described below); the specific interaction of a target polypeptide with another molecule, for example, a polypeptide that is activated by the target polypeptide or that inhibits the target polypeptide activity (e.g., by the two-hybrid assay); or the phosphorylation or glycosylation state of the target polypeptide. A compound, such as a dsRNA, that increases the level of the target polypeptide, mRNA encoding the target polypeptide, or reporter gene activity within a cell, a cell extract, or other experimental sample is a compound that stimulates or increases the biological activity of a target polypeptide. A compound, such as a dsRNA, that decreases the level of the target polypeptide, mRNA encoding the target polypeptide, or reporter gene activity within a cell, a cell extract, or other experimental sample is a compound that decreases the biological activity of a target polypeptide.

[0064] By "assaying" is meant analyzing the effect of a treatment, be it chemical or physical, administered to whole animals, cells, tissues, or molecules derived therefrom. The material being analyzed may be an animal, a cell, a tissue, a lysate or extract derived from a cell, or a molecule derived from a cell. The analysis may be, for example, for the purpose of detecting altered cell function, altered gene expression, altered endogenous RNA stability, altered polypeptide stability, altered polypeptide levels, or altered polypeptide biological activity. The means for analyzing may include, for example, antibody labeling, immunoprecipitation, phosphorylation assays, glycosylation assays, and methods known to those skilled in the art for detecting nucleic acids. In some embodiments, assaying is conducted under selective conditions.

[0065] By "modulates" is meant changing, either by a decrease or an increase. As used herein, desirably a nucleic acid sequence decreases the function of a cell, the expression of a target nucleic acid in a cell, or the biological activity of a target polypeptide in a cell by least 20%, more desirably by at least 30%, 40%, 50%, 60% or 75%, and most desirably by at least 90%. Also as used herein, desirably a nucleic acid sequence increases the function of a cell, the expression of a target nucleic acid in a cell, or the biological activity of a target

polypeptide in a cell by at least 1.5-fold to 2-fold, more desirably by at least 3-fold, and most desirably by at least 5-fold.

[0066] By "a decrease" is meant a lowering in the level of: a) protein (e.g., as measured by ELISA or Western blot analysis); b) reporter gene activity (e.g., as measured by reporter gene assay, for example, β -galactosidase, green fluorescent protein, or luciferase activity); c) mRNA (e.g., as measured by RT-PCR or Northern blot analysis relative to an internal control, such as a "housekeeping" gene product, for example, β -actin or glyceraldehyde 3-phosphate dehydrogenase (GAPDH)); or d) cell function, for example, as assayed by the number of apoptotic, mobile, growing, cell cycle arrested, invasive, differentiated, or dedifferentiated cells in a test sample. In all cases, the lowering is desirably by at least 20%, more desirably by at least 30%, 40%, 50%, 60%, 75%, and most desirably by at least 90%. As used herein, a decrease may be the direct or indirect result of PTGS.

[0067] By "an increase" is meant a rise in the level of: a) protein (e.g., as measured by ELISA or Western blot analysis); b) reporter gene activity (e.g., as measured by reporter gene assay, for example, β -galactosidase, green fluorescent protein, or luciferase activity); c) mRNA (e.g., as measured by RT-PCR or Northern blot analysis relative to an internal control, such as a "housekeeping" gene product, for example, β -actin or glyceraldehyde 3-phosphate dehydrogenase (GAPDH)); or d) cell function, for example, as assayed by the number of apoptotic, mobile, growing, cell cycle arrested, invasive, differentiated, or dedifferentiated cells in a test sample. Desirably, the increase is by at least 1.5-fold to 2-fold, more desirably by at least 3-fold, and most desirably by at least 5-fold. As used herein, an increase may be the indirect result of PTGS. For example, the double stranded RNA may inhibit the expression of a protein, such as a suppressor protein, that would otherwise inhibit the expression of another nucleic acid.

[0068] By "alteration in the level of gene expression" is meant a change in transcription, translation, or mRNA or protein stability such that the overall amount of a product of the gene, i.e., mRNA or polypeptide, is increased or decreased.

[0069] By "reporter gene" is meant any gene that encodes a product whose expression is detectable and/or able to be quantitated by immunological, chemical, biochemical, or biological assays. A reporter gene product may, for example, have one of the following attributes, without restriction: fluorescence (e.g., green fluorescent protein), enzymatic activity (e.g., β -galactosidase, luciferase, chloramphenicol acetyltransferase), toxicity (e.g., ricin A), or an ability to be specifically bound by an additional molecule (e.g., an unlabeled antibody, followed by a labelled secondary antibody, or biotin, or a detectably labelled antibody). It is understood that any engineered variants of reporter genes that are readily available to one skilled in the art, are also included, with-

out restriction, in the foregoing definition.

[0070] By "protein" or "polypeptide" or "polypeptide fragment" is meant any chain of more than two amino acids, regardless of post-translational modification (e.g., glycosylation or phosphorylation), constituting all or part of a naturally-occurring polypeptide or peptide, or constituting a non-naturally occurring polypeptide or peptide.

[0071] By "promoter" is meant a minimal sequence sufficient to direct transcription of a gene. Also included in this definition are those transcription control elements (e.g., enhancers) that are sufficient to render promoter-dependent gene expression controllable in a cell type-specific, tissue-specific, or temporal-specific manner, or that are inducible by external signals or agents; such elements, which are well-known to skilled artisans, may be found in a 5' or 3' region of a gene or within an intron. Desirably a promoter is operably linked to a nucleic acid sequence, for example, a cDNA or a gene in such a way as to permit expression of the nucleic acid sequence.

[0072] By "operably linked" is meant that a gene and one or more transcriptional regulatory sequences, e.g., a promoter or enhancer, are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequences.

[0073] By "expression vector" is meant a DNA construct that contains at least one promoter operably linked to a downstream gene or coding region (e.g., a cDNA or genomic DNA fragment that encodes a protein, optionally, operatively linked to sequence lying outside a coding region, an antisense RNA coding region, or RNA sequences lying outside a coding region). Transfection or transformation of the expression vector into a recipient cell allows the cell to express RNA encoded by the expression vector. An expression vector may be a genetically engineered plasmid, virus, or artificial chromosome derived from, for example, a bacteriophage, adenovirus, retrovirus, poxvirus, or herpesvirus.

[0074] By "transformation" or "transfection" is meant any method for introducing foreign molecules into a cell (e.g., a bacterial, yeast, fungal, algal, plant, insect, or animal cell, particularly a vertebrate or mammalian cell). The cell may be in an animal. Lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, viral or retroviral delivery, electroporation, and biolistic transformation are just a few of the transformation/transfection methods known to those skilled in the art. The RNA or RNA expression vector (DNA) may be naked RNA or DNA or local anesthetic complexed RNA or DNA (Pachuk *et al.*, *supra*). Other standard transformation/transfection methods and other RNA and/or DNA delivery agents (e.g., a cationic lipid, liposome, or bupivacaine) are described in WO 00/63364, filed April 19, 2000 (see, for example, pages 18-26). Commercially available kits can also be used to deliver RNA or DNA to a cell. For example, the Transmessenger Kit from Qiagen, an RNA kit

21

EP 1 229 134 A2

22

from Xeragon Inc., and an RNA kit from DNA Engine Inc. (Seattle, WA) can be used to introduce single or double stranded RNA into a cell.

[0075] By "transformed cell" or "transfected cell" is meant a cell (or a descendent of a cell) into which a nucleic acid molecule, for example, a double stranded RNA or double stranded expression vector has been introduced, by means of recombinant nucleic acid techniques. Such cells may be either stably or transiently transfected.

[0076] By "selective conditions" is meant conditions under which a specific cell or group of cells can be selected for. For example, the parameters of a fluorescence-activated cell sorter (FACS) can be modulated to identify a specific cell or group of cells. Cell panning, a technique known to those skilled in the art, is another method that employs selective conditions.

[0077] As use herein, by "optimized" is meant that a nucleic acid fragment is generated through inefficient first strand synthesis (e.g., reverse transcription (RT) and/or RT/second strand synthesis (RT-SSS) using Klenow or other enzymes and/or RT-PCR or PCR, to be of a particular length. Desirably the length of the nucleic acid fragment is less than a full length cDNA or is 100, 500, 600, or 1000 nucleotides in length. In other embodiments, the nucleic acid fragment is at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the nucleic acid fragment is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the nucleic acid fragment is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the nucleic acid fragment contains less than 50,000; 10,000; 5,000; or 2,000 nucleotides. Optimization of the length of a nucleic acid can be achieved during first strand or second strand synthesis of a desired nucleic acid by lowering Mg^{++} concentrations to no less than the nucleotide concentrations; by adding Mn^{++} to the reaction to achieve the desired size selection (e.g., by replacing Mg^{++} completely, or by adding Mn^{++} at varying concentrations along with Mg^{++}); by decreasing and/or limiting concentrations of dNTP(s) to effect the desired fragment size; by using various concentrations of ddNTP(s) along with standard or optimal concentrations of dNTP(s), to achieve varying ratios, to obtain the desired fragment size; by using limited and controlled exonuclease digestion of the fragment following RT, RT-SSS, RT-PCR, or PCR; or by a combination of any of these methods.

[0078] As used herein, by "sized selected" is meant that a nucleic acid of a particular size is selected for use in the construction of dsRNA expression libraries as de-

scribed herein. Desirably the size selected nucleic acid is less than a full length cDNA sequence or at least 100, 500, 600, or 1000 nucleotides in length. In other embodiments, the nucleic acid is at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the nucleic acid is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the nucleic acid is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the nucleic acid contains less than 50,000; 10,000; 5,000; or 2,000 nucleotides. For example, a nucleic acid may be size selected using size exclusion chromatography (e.g., as size exclusion Sepharose matrices) according to standard procedures (see, for example, Sambrook, Fritsch, and Maniatis, Molecular Cloning: A Laboratory Manual (2d ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

[0079] By "under conditions that inhibit or prevent an interferon response or a dsRNA stress response" is meant conditions that prevent or inhibit one or more interferon responses or cellular RNA stress responses involving cell toxicity, cell death, an anti-proliferative response, or a decreased ability of a dsRNA to carry out a PTGS event. These responses include, but are not limited to, interferon induction (both Type I and Type II), induction of one or more interferon stimulated genes, PKR activation, 2'5'-OAS activation, and any downstream cellular and/or organismal sequelae that result from the activation/induction of one or more of these responses. By "organismal sequelae" is meant any effect (s) in a whole animal, organ, or more locally (e.g., at a site of injection) caused by the stress response. Exemplary manifestations include elevated cytokine production, local inflammation, and necrosis. Desirably the conditions that inhibit these responses are such that not more than 95%, 90%, 80%, 75%, 60%, 40%, or 25%, and most desirably not more than 10% of the cells undergo cell toxicity, cell death, or a decreased ability to carry out a PTGS event, compared to a cell not exposed to such interferon response inhibiting conditions, all other conditions being equal (e.g., same cell type, same transformation with the same dsRNA expression library).

[0080] Apoptosis, interferon induction, 2'5' OAS activation/induction, PKR induction/activation, anti-proliferative responses, and cytopathic effects are all indicators for the RNA stress response pathway. Exemplary assays that can be used to measure the induction of an RNA stress response as described herein include a TUNEL assay to detect apoptotic cells, ELISA assays to detect the induction of alpha, beta and gamma interferon, ribosomal RNA fragmentation analysis to detect

activation of 2'5'OAS, measurement of phosphorylated eIF2 α as an indicator of PKR (protein kinase RNA inducible) activation, proliferation assays to detect changes in cellular proliferation, and microscopic analysis of cells to identify cellular cytopathic effects (see, e.g., Example 11). Desirably, the level of an interferon response or a dsRNA stress response in a cell transformed with a double stranded RNA or a double stranded RNA expression vector is less than 20, 10, 5, or 2-fold greater than the corresponding level in a mock-transfected control cell under the same conditions, as measured using one of the assays described herein. In other embodiments, the level of an interferon response or a dsRNA stress response in a cell transformed with a double stranded RNA or a double stranded RNA expression vector using the methods of the present invention is less than 500%, 200%, 100%, 50%, 25%, or 10% greater than the corresponding level in a corresponding transformed cell that is not exposed to such interferon response inhibiting conditions, all other conditions being equal. Desirably, the double stranded RNA does not induce a global inhibition of cellular transcription or translation.

[0081] By \square specifically hybridizes \square is meant a double stranded RNA that hybridizes to a target nucleic acid but does not substantially hybridize to other nucleic acids in a sample (e.g., a sample from a cell) that naturally includes the target nucleic acid, when assayed under denaturing conditions. In one embodiment, the amount of a target nucleic acid hybridized to, or associated with, the double stranded RNA, as measured using standard assays, is 2-fold, desirably 5-fold, more desirably 10-fold, and most desirably 50-fold greater than the amount of a control nucleic acid hybridized to, or associated with, the double stranded RNA.

[0082] By \square high stringency conditions \square is meant hybridization in 2X SSC at 40°C with a DNA probe length of at least 40 nucleotides. For other definitions of high stringency conditions, see F. Ausubel *et al.*, *Current Protocols in Molecular Biology*, pp. 6.3.1-6.3.6, John Wiley & Sons, New York, NY, 1994, hereby incorporated by reference.

[0083] Conditions and techniques that can be used to prevent an interferon response or dsRNA stress response during the screening methods of the present invention are described herein.

Brief Description of the Drawing

[0084] Fig. 1 is a schematic representation of a strategy to isolate clonally pure stable integrants that contain a single expression unit isolated from cells transfected with a double-stranded RNA encoding a cDNA library.

[0085] Fig. 2 is a schematic illustration of the production of effector RNAs in cells expressing PSA. The PSA expression cassette used to create the transient PSA expression cell line is depicted at the top of the figure. Expression of PSA is driven by the HCMV IE promoter

and the SV40 polyadenylation signal (pA). Only sequences 3' of the PSA initiation codon have been used in these vectors. The effector RNA expression cassettes are shown below the PSA expression cassette and are designed to express PSA sense RNA, PSA antisense RNA, and PSA dsRNA. Expression of the effector RNAs is under the control of the T7 promoter (T7p). Transcription from T7p is catalyzed by T7 RNA polymerase, which is supplied by co-transfecting a T7 RNA polymerase expression plasmid (not shown). Control effector RNA cassettes expressing irrelevant RNAs derived from the Herpes simplex virus glycoprotein D gene were included as controls. The 600 base pair sequence from the Herpes simplex gD gene is from Herpes Simplex virus 2 strain 12 and maps to the coding region downstream of the gD initiation codon.

[0086] Fig. 3 is a bar graph illustrating silencing of PSA expression by dsRNA. PSA levels in the supernates of transfected cells were determined by ELISA and are plotted as percent expression of the PSA untreated control. The PSA untreated control shown at the left is normalized to 100%. PSA levels in the supernates of cells transfected with the various effector PSA or control RNAs are shown by the shaded and open bars respectively. All data shown is from day two post-transfection. Data from later time points were similar to the day two time point.

[0087] Fig. 4 is a schematic illustration of the RNA stress response pathway, also known as the Type 1 interferon response.

Detailed Description of the Invention

[0088] Post-transcriptional gene silencing (PTGS) can be used as a tool to identify and validate specific unknown genes involved in cell function, gene expression, and polypeptide biological activity. Although the use of PTGS as a validation strategy is well documented in invertebrates and plants, its use in identification of genes that modulate cell function, gene expression, or polypeptide biological activity, as described below, is novel. Since novel genes are likely to be identified through the methods of the present invention, PTGS is developed for use in validation and to identify novel targets for use in therapies for diseases, for example, cancer, neurological disorders, obesity, leukemia, lymphomas, and other disorders of the blood or immune system.

[0089] The present invention features methods to identify unknown targets that result in the modulation of a particular phenotype, an alteration of gene expression in a cell, or an alteration in polypeptide biological activity in a cell, using either a library based screening approach or a non-library based approach to identify nucleic acids that induce gene silencing. The present invention also allows the determination of function of a given sequence. These methods involve the direct delivery of *in vitro* transcribed double stranded RNA (dsRNA), as well

25

EP 1 229 134 A2

26

as plasmid-based systems that direct the cell to make its own dsRNA. To avoid problems associated with transfection efficiency, plasmids are designed to contain a selectable marker to ensure the survival of only those cells that have taken up plasmid DNA. One group of plasmids directs the synthesis of dsRNA that is transcribed in the cytoplasm, while another group directs the synthesis of dsRNA that is transcribed in the nucleus.

Identification of genes by assaying for a modulation in cell function

[0090] Functional identification of novel genes can be accomplished through the use of a number of different assays. For example, cells may be assayed for cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, or the ability to support viral replication, as well as other cell functions known in the art. Methods for carrying out such functional assays are well known and are described, for example, in Platet and Garcia (Invasion Metastasis 18: 198-208, 1998-1999); Harper *et al.* (Neuroscience 88: 257-267, 1999); and Tomaselli *et al.* (J. Cell Biol. 105: 2347-2358, 1987), and are also described below.

[0091] Functional identification of nucleic acid sequences involved in modulating a particular cell function may be carried out by comparing cells transfected with a dsRNA to control cells that have not been transformed with a dsRNA or that have been mock-transfected, in a functional assay. A cell that has taken up sequences unrelated to a particular function will perform in the particular assay in a manner similar to the control cell. A cell experiencing PTGS of a gene involved in the particular function will exhibit an altered ability to perform in the functional assay compared to the control.

[0092] The percent modulation of a particular cell function that identifies a nucleic acid sequence that modulates the function of a cell will vary depending on the assay, phenotype, and the particular nucleic acid affected by PTGS. For each assay, the percent modulation can readily be determined by one skilled in the art, when used in conjunction with controls, as described herein. Desirably the modulation is at least 20%, more desirably at least 30%, 40%, 50%, 60%, 75%, and most desirably at least 90% compared to the control. An increase in the function of a cell can also be measured in terms of fold increase, where desirably, the increase is at least 1.5-fold to 5-fold compared to the control.

[0093] Alternatively, the function of a cell may be to affect the function, gene expression, or polypeptide biological activity of another cell, for example, a neighboring cell, a cell that is contacted with the cell in which a PTGS event occurs, or a cell that is contacted with media or other extracellular fluid that the cell in which a PTGS event occurs is contained in. For example, a cell experiencing PTGS of a gene may modulate cell motility, apoptosis, cell growth, cell invasion, vascularization,

cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, or the ability to support viral replication of a nearby cell, or a cell that is exposed to media or other extracellular fluid in which the transfected cell in which a PTGS event occurs was once contained. This can be tested by removing the media in which a cell experiencing a PTGS event is occurring and placing it on a separate cell or population of cells. If the function of the separate cell or population of cells is modulated, compared to a cell or population of cells receiving media obtained from cells that had been mock transfected, then one or more of the cells experiencing a PTGS event can affect the function of another cell. The identity of the nucleic acid sequence that causes the modulation can be identified with repeated rounds of selection.

[0094] In another method, a single cell experiencing a PTGS event can be placed in proximity of a cell or a population of cells that was not transfected with dsRNA, and the effect of this placement is evaluated for a modulation in the function of the cell or population of cells. If the function of the non-transfected cell or population of cells is modulated, compared to a cell or population of cells in proximity of a cell that was mock transfected, then the cell experiencing a PTGS event contains a nucleic acid sequence that can affect the function of another cell. This nucleic acid sequence can be identified using techniques described herein.

Identification of genes using differential gene expression

[0095] Differential gene expression analysis can be used to identify a nucleic acid sequence that modulates the expression of a target nucleic acid in a cell. Alterations in gene expression induced by gene silencing can be monitored in a cell into which a dsRNA has been introduced. For example, differential gene expression can be assayed by comparing nucleic acids expressed in cells into which dsRNA has been introduced to nucleic acids expressed in control cells that were not transfected with dsRNA or that were mock-transfected. Gene array technology can be used in order to simultaneously examine the expression levels of many different nucleic acids. Examples of methods for such expression analysis are described by Marrack *et al.* (Current Opinions in Immunology 12:206-209, 2000); Harkin (Oncologist 5:501-507, 2000); Pelizzari *et al.* (Nucleic Acids Res. 28: 4577-4581, 2000); and Marx (Science 289:1670-1672, 2000).

Identification of genes by assaying polypeptide biological activity

[0096] Novel nucleic acid sequences that modulate the biological activity of a target polypeptide can also be identified by examining polypeptide biological activity. Various polypeptide biological activities can be evaluat-

ed to identify novel genes according to the methods of the invention. For example, the expression of a target polypeptide(s) may be examined. Alternatively, the interaction between a target polypeptide(s) and another molecule(s), for example, another polypeptide or a nucleic acid may be assayed. Phosphorylation or glycosylation of a target polypeptide(s) may also be assessed, using standard methods known to those skilled in the art.

[0097] Identification of nucleic acid sequences involved in modulating the biological activity of a target polypeptide may be carried out by comparing the polypeptide biological activity of a cell transfected with a dsRNA to a control cell that has not been transfected with a dsRNA or that has been mock-transfected. A cell that has taken up sequences unrelated to a particular polypeptide biological activity will perform in the particular assay in a manner similar to the control cell. A cell experiencing PTGS of a gene involved in the particular polypeptide biological activity will exhibit an altered ability to perform in the biological assay, compared to the control.

Insertion of single units into the chromosome and generation of a cell line containing a single dsRNA expression library integrant

[0098] The present invention involves the generation of a target cell line in which the dsRNA expression library is subsequently introduced. Through the use of site-specific recombination, single integrants of dsRNA expression cassettes are generated at the same locus of all cells in the target cell line, allowing uniform expression of the dsRNA in all of the integrants. A dsRNA expression library derived from various cell lines is used to create a representative library of stably integrated cells, each cell within the target cell line containing a single integrant. Cre/lox, Lambda-Cro repressor, and Flp recombinase systems or retroviruses are used to generate these singular integrants of dsRNA expression cassettes in the target cell line (Sato *et al.*, J. Virol. 74: 10631-10638, 2000; Trinh *et al.*, J. Immunol. Methods 244:185-193, 2000; Serov *et al.*, An. Acad. Bras. Cienc. 72:389-398, 2000; Grez *et al.*, Stem Cells. 16:235-243, 1998; Habu *et al.*, Nucleic Acids Symp. Ser. 42:295-296, 1999; Haren *et al.*, Annu. Rev. Microbiol. 53:245-281, 1999; Baer *et al.*, Biochemistry 39:7041-7049, 2000; Follenzi *et al.*, Nat. Genet. 25:217-222, 2000; Hindmarsh *et al.*, Microbiol. Mol. Biol. Rev. 63:836-843, 1999; Darquet *et al.*, Gene Ther. 6:209-218, 1999; Darquet *et al.*, Gene Ther. 6:209-218, 1999; Yu *et al.*, Gene 223:77-81, 1998; Darquet *et al.*, Gene Ther. 4: 1341-1349, 1997; and Koch *et al.*, Gene 249:135-144, 2000). These systems are used singularly to generate singular insertion clones, and also in combination.

[0099] The following exemplary sequence specific integrative systems use short target sequences that allow targeted recombination to be achieved using specific

proteins: FLP recombinase, bacteriophage Lambda integrase, HIV integrase, and pilin recombinase of *Salmonella* (Seng *et al.* Construction of a Flp "exchange cassette" contained vector and gene targeting in mouse ES cell] A book chapter PUBMED entry 11797223 - Sheng Wu Gong Cheng Xue Bao. 2001 Sep;17(5):566-9., Liu *et al.*, Nat Genet. 2001 Jan 1;30(1):66-72., Awatramani *et al.*, Nat Genet. 2001 Nov;29(3):257-9., Heldmann and Lehner, Dev Genes Evol. 2001 Sep;211(8-9):458-65, Schaft *et al.*, Genesis. 2001 Sep;31(1):6-10, Van Duyne, Annu Rev Biophys Biomol Struct. 2001;30: 87-104., Lorbach *et al.*, J Mol Biol. 2000 Mar 10;296(5): 1175-81., Darquet *et al.*, Gene Ther. 1999 Feb;6(2): 209-18., Bushman and Miller, J Virol. 1997 Jan;71(1): 458-64., Fulks *et al.*, J Bacteriol. 1990 Jan;172(1): 310-6). A singular integrant is produced by randomly inserting the specific sequence (e.g., loxP in the cre recombinase system) and selecting or identifying the cell that contains a singular integrant that supports maximal expression. For example, integrants that show maximal expression following random integration can be identified through the use of reporter gene sequences associated with the integrated sequence. The cell can be used to specifically insert the expression cassette into the site that contains the target sequence using the specific recombinase, and possibly also remove the expression cassette that was originally placed to identify the maximally expressing chromosomal location. A skilled artisan can also produce singular integrants using retroviral vectors, which integrate randomly and singularly into the eukaryotic genome. In particular, singular integrants can be produced by inserting retroviral vectors that have been engineered to contain the desired expression cassette into a naive cell and selecting for the chromosomal location that results in maximal expression (Michael *et al.*, EMBO Journal, vol 20: pages 2224-2235, 2001; Reik and Murrell., Nature, vol. 405, page 408-409, 2000; Berger *et al.*, Molecular Cell, vol. 8, pages 263-268). One may also produce a singular integrant by cotransfecting the bacterial RecA protein with or without nuclear localization signal along with sequences that are homologous to the target sequence (e.g., a target endogenous sequence or integrated transgene sequence). Alternatively, a nucleic acid sequence that encodes a RecA protein with nuclear localization signals can be cotransfected (Shibata *et al.*, Proc Natl Acad Sci U S A. 2001 Jul 17;98(15):8425-32. Review., Muylers *et al.*, Trends Biochem Sci. 2001 May;26(5): 325-31., Paul *et al.*, Mutat Res. 2001 Jun 5;486(1):11-9., Shcherbakova *et al.*, Mutat Res. 2000 Feb 16;459(1): 65-71., Lantsov. Mol Biol (Mosk). 1994 May-Jun;28(3): 485-95).

[0100] An example utilizing such methods is detailed below.

Creation of the target cell line

[0101] Target cell lines are the same cell lines as the

ones from which the dsRNA expression libraries will be derived. Target cells are created by transfecting the selected cell line with a bicistronic plasmid expressing a selectable marker, such as G418 and the reporter gene GFP. The plasmid also bears a *loxP* site. Plasmids integrate randomly into the chromosome through the process of illegitimate recombination at a frequency of 10^{-4} . Following transfection, cells containing integrants are selected by culturing the cells in the presence of G418 at a concentration determined earlier in a kill curve analysis. About a dozen G418-resistant colonies are expanded and relative GFP expression levels are determined using flow cytometry. DNA from the cells is analyzed by Southern blot analysis to determine integrant copy number. Several single copy integrants exhibiting the highest GFP expression levels are then selected as the target cell lines. GFP expression is monitored because dsRNA encoding templates are then integrated into the loci containing the *loxP*, GFP, and G418 cassettes in a site-specific fashion, and it is important to ensure that these loci are transcriptionally active. Since cells are selected on the basis of G418 resistance and GFP expression, integration of the plasmid DNA can occur at the *loxP* site, destroying its function. Several cell lines are therefore chosen to reasonably ensure that at least one integrant has an intact *loxP* site.

Double stranded RNA expression library construction and site-specific recombination into the target cell line

[0102] A cDNA library or a randomized library is constructed from RNA isolated from selected cell lines. cDNAs or randomized nucleic acids in the size range of at least 100 to 1000 nucleotides, for example, 500 to 600 nucleotides are optimized during synthesis or are size-selected prior to cloning. In other embodiments, the nucleic acids are at least 10, 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In yet other embodiments, the number of nucleotides in the nucleic acids is between 5-100 nucleotides, 15-100 nucleotides, 20-95 nucleotides, 25-90 nucleotides, 35-85 nucleotides, 45-80 nucleotides, 50-75 nucleotides, or 55-70 nucleotides, inclusive. In still other embodiments, the number of nucleotides in the nucleic acids is contained in one of the following ranges: 5-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-35 nucleotides, 35-45 nucleotides, 45-60 nucleotides, 60-70 nucleotides, 70-80 nucleotides, 80-90 nucleotides, or 90-100 nucleotides, inclusive. In other embodiments, the nucleic acid contains less than 50,000; 10,000; 5,000; or 2,000 nucleotides. Each cDNA or randomized nucleic acid is then cloned into a plasmid vector as a dsRNA transcription cassette flanked by two convergent promoters (such as T7 promoters as described herein). The promoters are transcriptionally regulated such that they are off until induced, for example, using a tet ON/OFF system (Forster *et al.*, *Nucleic Acids Res.* 27:7708-710, 1999; Liu *et al.*, *Biotechniques* 24:624-628, 6,30-632, 1998; and Gatz,

Methods Cell Biol. 50:411-424, 1995). The plasmid also contains the hygromycin resistance gene and an inverted *loxP* site. The cDNA plasmid library or randomized plasmid library is then co-transfected into the target cell line with a plasmid expressing Cre recombinase, which catalyzes site-specific recombination of the transfected cDNA plasmid or randomized nucleic acid plasmid at the inverted *loxP* site into the chromosomal locus containing the GFP gene and *loxP* site (see Fig. 1). The use of the Cre/*lox* system allows the efficient integration of a plasmid into the chromosome (every transfected cell is predicted to undergo a plasmid integration event). Other site-specific recombination strategies can also be utilized. This results in having every integration to occur at the same site, thereby obviating potential problems with loci dependent expression.

[0103] Two days following transfection, cells are incubated in the presence of hygromycin to kill untransfected cells and to select for stable integrants. Transcription of dsRNA is induced, and selected cells are assayed for an alteration in cell function, the biological activity of a target polypeptide, or differential gene expression. Cells expressing dsRNA corresponding to a target nucleic acid exhibit an altered function, for example, increased or decreased cell invasion, motility, apoptosis, growth, differentiation, dedifferentiation, or regeneration, or the ability of the cell to support viral replication. Cells exhibiting altered function are then expanded and the sequence of the integrant is determined. Targets are identified and validated using dsRNA specific for the identified target, or other non-PTGS mediated methods, for example antisense technology.

[0104] The regulated transcription system of the present invention provides an advantage to other double stranded expression systems. Following transfection of the dsRNA library, cells contain hundreds to thousands of dsRNA expression cassettes, with concomitant expression of that many expression cassettes. In the dsRNA expression system of the present invention, dsRNA expression cassettes contained within the expression vector integrate into the chromosome of the transfected cell. As described in detail below, every transfected cell integrates one of the double stranded expression cassettes. Desirably no transcription occurs until the episomal (non-integrated) expression vectors are diluted out of the cell such that not more than 5 episomal vectors remain in the cell. Most desirably no transcription occurs until all of the episomal (non-integrated) expression vectors are diluted out of the cell and only the integrated expression cassette remains (a process usually taking about two to several weeks of cell culture). At this time transcription is induced, allowing dsRNA to be expressed in the cells. This method ensures that only one species of dsRNA is expressed per cell, as opposed to other methods that express hundreds to thousands of double stranded species. The use of the above-described system results in the loss of all but one expression cassette, which in turn, permits the rapid

31

EP 1 229 134 A2

32

screening of libraries without requiring screening multiple pools of libraries to identify the target gene.

Non-library approaches for the identification of a nucleic acid sequence that modulates cell function, gene expression in a cell, or the biological activity of a target polypeptide in a cell through PTGS

[0105] Nucleic acid sequences that modulate cell function, gene expression in a cell, or the biological activity of a target polypeptide in a cell may also be identified using non-library based approaches involving PTGS. For example, a single known nucleic acid sequence encoding a polypeptide with unknown function or a single nucleic acid fragment of unknown sequence and/or function can be made into a double stranded RNA molecule. This dsRNA is then transfected into a desired cell type and the cell is assayed for modulations in cell function, gene expression of a target nucleic acid in the cell, or the biological activity of a target polypeptide in the cell, using methods described herein. A modulation in cell function, gene expression in the cell, or the biological activity of a target polypeptide in the cell identifies the nucleic acid of the dsRNA as a nucleic acid that modulates the specific cell function, gene expression, or the biological activity of a target polypeptide. As a single dsRNA species is transfected into the cells, the nucleic acid sequence responsible for the modulation is readily identified. This non-library based approach to nucleic acid identification is desirably used under conditions that inhibit an interferon response or dsRNA stress response. Such conditions are described in detail herein.

[0106] The discovery of novel genes through the methods of the present invention may lead to the generation of novel therapeutics. For example, genes that decrease cell invasion may be used as targets for drug development, such as for the development of cytostatic therapeutics for use in the treatment of cancer. Development of such therapeutics is important because currently available cytotoxic anticancer agents are also toxic for normal rapidly dividing cells. In contrast, a cytostatic agent may only need to check metastatic processes, and by inference, slow cell growth, in order to stabilize the disease. In another example, genes that increase neuronal regeneration may be used to develop therapeutics for the treatment, prevention, or control of a number of neurological diseases, including Alzheimer's disease and Parkinson's disease. Genes that are involved in the ability to support viral replication and be used as targets in anti-viral therapies. Such therapies may be used to treat, prevent, or control viral diseases involving human immunodeficiency virus (HIV), hepatitis C virus (HCV), hepatitis B virus (HBV), and human papillomavirus (HPV). The efficacies of therapeutics targeting the genes identified according to the present invention can be further tested in cell culture assays, as well as in animal models.

The use of vertebrate or mammalian cell lines for identification of nucleic acid sequences that modulate cell function, expression of a target nucleic acid or biological activity of a target polypeptide

[0107] While the use of the present invention is not limited to vertebrate or mammalian cells, such cells can be used to carry out the nucleic acid identification methods described herein. Desirably the vertebrate or mammalian cells used to carry out the present invention are cells that have been cultured for only a small number of passages (e.g., less than 30 passages of a cell line that has been obtained directly from American Type Culture Collection), or are primary cells. In addition, vertebrate or mammalian cells can be used to carry out the present invention when the dsRNA being transfected into the cell is not complexed with cationic lipids.

[0108] The following examples are to illustrate the invention. They are not meant to limit the invention in any way. For example, it is noted that any of the following examples can be used with double stranded RNAs of any length. The methods of the present invention can be readily adapted by one skilled in the art to utilize double stranded RNAs of any desired length.

Example 1: Design and delivery of vectors for intracellular synthesis of dsRNA for library based screening approaches to nucleic acid identification using PTGS

[0109] PTGS is induced when dsRNA is made intracellularly. The library based screening approaches to nucleic acid identification through PTGS may require that dsRNA reside in certain cellular compartments in order to exert its effect. Therefore, expression plasmids that transcribe dsRNA in the cytoplasm and in the nucleus are utilized. There are two classes of nuclear transcription vectors: one that is designed to express polyadenylated dsRNA (for example, a vector containing an RNA polymerase II promoter and a poly A site) and one that expresses non-adenylated dsRNA (for example, a vector containing an RNA polymerase II promoter and no poly A site, or a vector containing a T7 promoter). Different cellular distributions are predicted for the two species of RNA; both vectors are transcribed in the nucleus, but the ultimate destinations of the RNA species are different intracellular locations. Intracellular transcription may also utilize bacteriophage T7 and SP6 RNA polymerase, which may be designed to transcribe in the cytoplasm or in the nucleus. Alternatively, Qbeta replicase RNA-dependent RNA polymerase may be used to amplify dsRNA. Viral RNA polymerases, either DNA and RNA dependent, may also be used. Alternatively, dsRNA replicating polymerases can be used. Cellular polymerases such as RNA Polymerase I, II, or III or mitochondrial RNA polymerase may also be utilized. Both the cytoplasmic and nuclear transcription vectors contain an antibiotic resistance gene to enable selection

33

EP 1 229 134 A2

34

of cells that have taken up the plasmid. Cloning strategies employ chain reaction cloning (CRC), a one-step method for directional ligation of multiple fragments (Pachuk *et al.*, Gene 243:19-25, 2000). Briefly, the ligations utilize bridge oligonucleotides to align the DNA fragments in a particular order and ligation is catalyzed by a heat-stable DNA ligase, such as Ampligase, available from Epicentre.

Inducible or repressible transcription vectors for the generation of a dsRNA expression library

[0110] If desired, inducible and repressible transcription systems can be used to control the timing of the synthesis of dsRNA. Inducible and repressible regulatory systems involve the use of promoter elements that contain sequences that bind prokaryotic or eukaryotic transcription factors upstream of the sequence encoding dsRNA. In addition, these factors also carry protein domains that transactivate or transrepress the RNA polymerase II. The regulatory system also has the ability to bind a small molecule (*e.g.*, a coinducer or a corepressor). The binding of the small molecule to the regulatory protein molecule (*e.g.*, a transcription factor) results in either increased or decreased affinity for the sequence element. Both inducible and repressible systems can be developed using any of the inducer/transcription factor combinations by positioning the binding site appropriately with respect to the promoter sequence. Examples of previously described inducible/repressible systems include *lacI*, *ara*, Steroid-RU486, and ecdysone - Rheogene, *Lac* (Cronin *et al.* *Genes & Development* 15: 1506-1517, 2001), *ara* (Khlebnikov *et al.*, J Bacteriol. 2000 Dec;182(24):7029-34), ecdysone (Rheogene, www.rheogene.com), RU48 (steroid, Wang XJ, Liefer KM, Tsal S, O'Malley BW, Roop DR., Proc Natl Acad Sci U S A. 1999 Jul 20;96(15):8483-8), *tet* promoter (Rendal *et al.*, Hum Gene Ther. 2002 Jan;13(2): 335-42. and Lamartina *et al.*, Hum Gene Ther. 2002 Jan; 13(2):199-210), or a promoter disclosed in WO 00/63364, filed April 19, 2000.

Nuclear transcription vectors for the generation of a nuclear dsRNA expression library

[0111] Nuclear transcription vectors for use in library based screening approaches to identify nucleic acids that modulate cell function, gene expression, or the biological activity of a target polypeptide are designed such that the target sequence is flanked on one end by an RNA polymerase II promoter (for example, the HCMV-IE promoter) and on the other end by a different RNA polymerase II promoter (for example, the SCMV promoter). Other promoters that can be used include other RNA polymerase II promoters, an RNA polymerase I promoter, an RNA polymerase III promoter, a mitochondrial RNA polymerase promoter, or a T7 or SP6 promoter in the presence of T7 or SP6 RNA polymerase,

respectively, containing a nuclear localization signal. Bacteriophage or viral promoters may also be used. The promoters are regulated transcriptionally (for example, using a tet ON/OFF system (Forster *et al.*, *supra*; Liu *et al.*, *supra*; and Gatz, *supra*) such that they are only active in either the presence of a transcription-inducing agent or upon the removal of a repressor. A single chromosomal integrant is selected for, and transcription is induced in the cell to produce the nuclear dsRNA.

[0112] Those vectors containing a promoter recognized by RNA Pol I, RNA Pol II, or a viral promoter in conjunction with co-expressed proteins that recognize the viral promoter, may also contain optional sequences located between each promoter and the inserted cDNA. These sequences are transcribed and are designed to prevent the possible translation of a transcribed cDNA. For example, the transcribed RNA is synthesized to contain a stable stem-loop structure at the 5' end to impede ribosome scanning. Alternatively, the exact sequence is irrelevant as long as the length of the sequence is sufficient to be detrimental to translation initiation (*e.g.*, the sequence is 200 nucleotides or longer). The RNA sequences can optionally have sequences that allow polyA addition, intronic sequences, an HIV REV binding sequence, Mason-Pfizer monkey virus constitutive transport element (CTE) (U.S. 5,880,276, filed April 25, 1996), and/or self splicing intronic sequences.

[0113] To generate dsRNA, two promoters can be placed on either side of the target sequence, such that the direction of transcription from each promoter is opposing each other. Alternatively, two plasmids can be cotransfected. One of the plasmids is designed to transcribe one strand of the target sequence while the other is designed to transcribe the other strand. Single promoter constructs may be developed such that two units of the target sequence are transcribed in tandem, such that the second unit is in the reverse orientation with respect to the other. Alternate strategies include the use of filler sequences between the tandem target sequences.

Cytoplasmic transcription vectors for the generation of a cytoplasmic dsRNA expression library

[0114] Cytoplasmic transcription vectors for use in library based screening approaches to identifying nucleic acids that modulate cell function, gene expression, or the biological activity of a target polypeptide in a cell using PTGS are made according to the following method. This approach involves the transcription of a single stranded RNA template (derived from a library) in the nucleus, which is then transported into the cytoplasm where it serves as a template for the transcription of dsRNA molecules. The DNA encoding the ssRNA is integrated at a single site in the target cell line as described for the nuclear RNA expression library, thereby ensuring the synthesis of only one species of dsRNA in a cell, each cell expressing a different dsRNA species.

[0115] A desirable approach is to use endogenous polymerases such as the mitochondrial polymerase in animal cells or mitochondrial and chloroplast polymerases in plant cells for cytoplasmic and mitochondrial (e.g., chloroplast) expression to make dsRNA in the cytoplasm. These vectors are formed by designing expression constructs that contain mitochondrial or chloroplast promoters upstream of the target sequence. As described above for nuclear transcription vectors, dsRNA can be generated using two such promoters placed on either side of the target sequence, such that the direction of transcription from each promoter is opposing each other. Alternatively, two plasmids can be cotransfected. One of the plasmids is designed to transcribe one strand of the target sequence while the other is designed to transcribe the other strand. Single promoter constructs may be developed such that two units of the target sequence are transcribed in tandem, such that the second unit is in the reverse orientation with respect to the other. Alternate strategies include the use of filler sequences between the tandem target sequences.

[0116] Alternatively, cytoplasmic expression of dsRNA for use in library based screening approaches is achieved by a single subgenomic promoter opposite in orientation with respect to the nuclear promoter. The nuclear promoter generates one RNA strand that is transported into the cytoplasm, and the singular subgenomic promoter at the 3' end of the transcript is sufficient to generate its antisense copy by an RNA dependent RNA polymerase to result in a cytoplasmic dsRNA species.

Target cell line development for use with cytoplasmic dsRNA expression libraries

[0117] The target cell line, using the vector containing the G418 cassette, GFP, and *loxP* site is designed as described above.

Development of a cytoplasmic dsRNA expression library

[0118] Double stranded RNA expression libraries are generated by inserting cDNA or randomized sequences (as described herein) into an expression vector containing a single nuclear promoter (RNA polymerase I, RNA polymerase II, or RNA polymerase III), which allows transcription of the insert sequence. It is desirable that this nuclear promoter activity is regulated transcriptionally (for example, using a tet ON/OFF system described, for example, by Forster *et al.*, *supra*; Liu *et al.*, *supra*; and Gatz, *supra*), such that the promoters are only active in either the presence of a transcription-inducing agent or upon the removal of a repressor. This ensures that transcription is not induced until episomal copies of the vector(s) are diluted out. Vectors also contain a selectable marker, such as the hygromycin resistance gene, and a *loxP* site. The expression vectors are integrated into the target cell line by methods previously de-

scribed in this application using Cre recombinase (other site-specific recombinative strategies can be employed, as described previously).

[0119] At two days post-transfection, cells are subjected to hygromycin selection using concentrations established in kill curve assays. Surviving cells are cultured in hygromycin to select for cells bearing integrated vectors and to dilute out episomal copies of the vector (s). At this point transcription is induced, and a single stranded RNA (ssRNA) species derived from the insert sequence is transcribed in the nucleus from the nuclear promoter in the inserted vector. The insert is designed such that the insert sequences in the transcript are flanked by bidirectional promoters of RNA bacteriophages (for example, Qbeta or MS2, RNA dependent RNA polymerase promoters) or cytoplasmic viral RNA-dependent RNA polymerase promoter sequences (for example, those of Sindbis or VEEV subgenomic promoters). The nuclear transcript is translocated to the cytoplasm where it acts as a template for dsRNA by an RNA dependent RNA polymerase, which may be provided through co-transfection of a vector that encodes an RNA-dependent RNA polymerase. Alternatively, an integrated copy of the polymerase may be used.

Example 2: Generation of templates for *in vitro* transcription of dsRNA for non-library based approaches for identification of nucleic acids using PTGS

[0120] Nucleic acid sequences that modulate cell function, gene expression in a cell, or the biological activity of a target polypeptide in a cell may also be identified using non-library based approaches involving PTGS. A single known nucleic acid sequence encoding a polypeptide with unknown function or a single nucleic acid fragment of unknown sequence and/or function can be made into a double stranded RNA molecule. This dsRNA is then transfected into a desired cell type and assayed for modulations in cell function, gene expression in the cell, or the biological activity of a target polypeptide in the cell, using methods described herein. A modulation in cell function, gene expression in the cell, or the biological activity of a target polypeptide in the cell identifies the nucleic acid of the dsRNA as a nucleic acid that modulates the specific cell function, gene expression, or the biological activity of a target polypeptide. This non-library based approach to nucleic acid identification is desirably used under conditions that inhibit an interferon response or dsRNA stress response. Such conditions are described in detail below.

[0121] Nucleic acid fragments generated, for example, by PCR or restriction endonuclease digestion, encoding the respective target sequences were used as templates for *in vitro* transcription reactions. PCR fragments are superior to plasmid templates for the synthesis of discrete sized RNA molecules. The PCR fragments encoded at least 20-50 or 100 to 1000, for exam-

ple, 500 to 600 nucleotides (nts) of the target sequence and were derived from the target mRNA. Known target sequences were obtained from GenBank and or other DNA sequence databases. Target sequences were also obtained from cellular RNAs that were generated into cDNAs to create a number of different dsRNA molecules. Accordingly, it is possible that the sequence and/or function of the target sequence was not known at the time the dsRNA was generated.

[0122] Templates for sense target RNAs were generated by placing the bacteriophage T7 promoter at the 5' end of the target coding strand while antisense RNA templates contained the T7 promoter at the 5' end of the non-coding strand. This was achieved by encoding the T7 promoter at the 5' ends of the respective PCR primers. Alternatively SP6 promoters, or a combination of SP6 and T7 promoters may be used.

[0123] PCR was performed by conventional methods. The use of both PCR templates in equimolar amounts in an *in vitro* transcription reaction resulted in primarily dsRNA. The use of two separate fragments has been found to be superior to the use of one PCR fragment containing two T7 promoters, one located at each end of the target sequence, presumably due to transcription interference that occurs during transcription of the dual promoter template. Following PCR amplification, the DNA was subjected to Proteinase K digestion and phenol-chloroform extraction to remove contaminating RNases. Following ethanol precipitation, the DNA was resuspended in RNase-free water at a concentration of 1 to 3 µg/µl.

[0124] As an alternative to phenol-chloroform extraction, DNA can be purified in the absence of phenol using standard methods such as those described by Li *et al.* (WO 00/44914, filed January 28, 2000). Alternatively, DNA that is extracted with phenol and/or chloroform can be purified to reduce or eliminate the amount of phenol and/or chloroform. For example, standard column chromatography can be used to purify the DNA (WO 00/44914, filed January 28, 2000).

Example 3: *In vitro* RNA transcription and RNA analysis

[0125] *In vitro* transcription reactions are carried out using the Riboprobe Kit (Promega Corp.), according to the manufacturer's directions. The template DNA is as described above. Following synthesis, the RNA is treated with RQ1 DNase (Promega Corp.) to remove template DNA. The RNA is then treated with Proteinase K and extracted with phenol-chloroform to remove contaminating RNases. The RNA is ethanol precipitated, washed with 70% ethanol, and resuspended in RNase-free water. Aliquots of RNA are removed for analysis and the RNA solution is flash frozen by incubating in an ethanol-dry ice bath. The RNA is stored at -80°C.

[0126] As an alternative to phenol-chloroform extraction, RNA can be purified in the absence of phenol using standard methods such as those described by Li *et al.*

(WO 00/44914, filed January 28, 2000). Alternatively, RNA that is extracted with phenol and/or chloroform can be purified to reduce or eliminate the amount of phenol and/or chloroform. For example, standard column chromatography can be used to purify the RNA (WO 00/44914, filed January 28, 2000).

[0127] dsRNA is made by combining equimolar amounts of PCR fragments encoding antisense RNA and sense RNA, as described above, in the transcription reaction. Single stranded antisense or sense RNA is made by using a single species of PCR fragment in the reaction. The RNA concentration is determined by spectrophotometric analysis, and RNA quality is assessed by denaturing gel electrophoresis and by digestion with RNase T1, which degrades single stranded RNA.

[0128] An mRNA library is produced using Qbeta bacteriophage, by ligating the mRNAs to the flank sequences that are required for Qbeta replicase function (Qbeta flank or Qbeta flank plus P1), using RNA ligase. The ligated RNAs are then transformed into bacteria that express Qbeta replicase and the coat protein. Single plaques are then inoculated into fresh bacteria. All plaques are expected to carry transgene sequences. Each plaque is grown in larger quantities in bacteria that produce the Qbeta polymerase, and RNA is isolated from the bacteriophage particles. Alternatively, if the Qbeta flank plus P1 is used to generate the library (e.g., P1=MS2, VEEV, or Sindbis promoter sequences), these vectors can be used to carry out the *in vitro* transcription along with the cognate polymerase. The *in vitro* made dsRNA is then used to transfect cells.

RNA delivery

[0129] *In vitro* made dsRNA is directly added to the cell culture medium at concentrations ranging from 50 µg/ml to 500 µg/ml. Uptake of dsRNA is also facilitated by electroporation using those conditions required for DNA uptake by the desired cell type. RNA uptake is also mediated by lipofection using any of a variety of commercially available and proprietary cationic lipids, DE-AE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, viral or retroviral delivery, or biolistic transformation. The RNA is naked RNA or a local anesthetic RNA complex. Modulation of cell function, gene expression, or polypeptide biological activity is then assessed in the transfected cells.

[0130] Some dsRNA sequences, possibly in certain cell types and through certain delivery methods, may result in an interferon response. During the screening methods of the present invention, induction of an interferon response is not desired, as this would lead to cell death and possibly to the prevention of gene silencing.

[0131] One of the components of an interferon response is the induction of the interferon-induced protein kinase PKR. Suppression of the interferon response and/or the PKR response is desired in the target cells.

The dsRNA delivery methods described herein are performed such that an interferon response or dsRNA stress response is not included. It is recognized, however, that certain conditions might present with an induction of the interferon response. To prevent such a response, a number of other strategies may be employed with any of the above described screening methods to identify a nucleic acid that modulates cell function, gene expression, or the polypeptide biological activity of a cell, as described herein.

[0132] To prevent an interferon response, interferon and PKR responses are silenced in the target cells using a dsRNA species directed against the mRNAs that encode proteins involved in the response. Desirably interferon response promoters are silenced using dsRNA. Alternatively, the expression of proteins that bind the interferon response element is abolished using dsRNA techniques.

[0133] In an alternative strategy, interferon and PKR knockout cell lines are created through approaches utilizing expression cassettes that encode an antisense RNA and ribozymes directed to the cellular mRNAs that encode the proteins involved in the response. Knockout cells are created by standard gene knockout technologies using homologous recombination to alter target sequences, using homologous DNA alone, or as complexes of RecA protein and single stranded DNA homologous to the target sequence(s). Interferon response element (IRE) sequences, sequences that encode transcription factors that bind IRE sequences, the promoter and/or gene sequences that encode proteins in the PKR and Interferon response pathways are molecules that are targeted for knockout.

[0134] If desired, proteins involved in gene silencing such as Dicer or Argonaut can be overexpressed or activated to increase the amount of inhibition of gene expression (Beach *et al.*, WO 01/68836, filed March 16, 2001).

Example 4: Cytoplasmic transcription vectors for non-library based approaches to nucleic acid identification using PTGS

[0135] Double stranded RNA molecules for use in non-library based methods for the identification of nucleic acids that modulate cell function, gene expression of a target nucleic acid, or target polypeptide biological activity in a cell can also be generated through the use of cytoplasmic transcription vectors. Such vectors are generated as now described.

[0136] The PCR fragments generated for *in vitro* transcription templates, as described above, are inserted into a cloning vector containing one T7 promoter located just outside the polylinker region. Such a vector is pZERO blunt (Promega Corp.). The PCR fragment is cloned into a restriction site in the polylinker in such a way that the fragment's T7 promoter is distal to the vector's promoter. The resulting vector contains the target se-

quence flanked by two T7 promoters; transcription from this vector occurs in converging directions. Convergent transcription is desired for these intracellular vectors, due to the uncertainty of getting sense and antisense vectors into the same cell in high enough and roughly equivalent amounts. In addition, the local concentration of antisense and sense RNAs with respect to each other is high enough to enable dsRNA formation when the dual promoter construct is used.

[0137] A hygromycin resistance cassette is cloned into the pZERO blunt vector as well. The hygromycin resistance cassette contains the hygromycin resistance gene under the control of the Herpes Simplex Virus (HSV) thymidine kinase promoter and the SV40 polyadenylation signal. The cassette is in a plasmid vector and is flanked at both ends by a polylinker region enabling ease of removal and subsequent cloning. Hygromycin selection was chosen because of the rapidity of death induced by hygromycin as well as extensive in-house experience with hygromycin selection. Alternatively, other selection methods known to those skilled in the art may be used.

[0138] The vectors are transfected into the desired cells using standard transformation or transfection techniques described herein, and the cells are assayed for the ability of the dsRNA molecules encoded by the vectors to modulate cell function, gene expression of a target nucleic acid, or the biological activity of a target polypeptide, as described herein.

Example 5: Analysis of RNA from transfected cells

[0139] Regardless of whether a library based screening approach or a non-library based approach was used to identify nucleic acid sequences, in order to measure the level of dsRNA effector molecule within the cell, as well as the amount of target mRNA within the cell, a two-step reverse transcription PCR reaction was performed with the ABI PRISM™ 7700 Sequence Detection System. Total RNA was extracted from cells transfected with dsRNA or a plasmid from a dsRNA expression library using Trizol and DNase. Two to three different cDNA synthesis reactions were performed per sample; one for human GAPDH (a housekeeping gene that should be unaffected by the effector dsRNA), one for the target mRNA, and/or one for the sense strand of the expected dsRNA molecule (effector molecule). Prior to cDNA synthesis of dsRNA sense strands, the RNA sample was treated with T1 RNase. The cDNA reactions were performed in separate tubes using 200 ng of total RNA and primers specific for the relevant RNAs. The cDNA products of these reactions were used as templates for subsequent PCR reactions to amplify GAPDH, the target cDNA, and/or the sense strand copied from the dsRNA. All RNA was quantified relative to the internal control, GAPDH.

41

EP 1 229 134 A2

42

Example 6: Target sequence identification

[0140] To identify the target sequence affected by a dsRNA, using any of the above-described methods, DNA is extracted from expanded cell lines (or from the transfected cells if using a non-integrating dsRNA system) according to methods well known to the skilled artisan. The dsRNA encoding sequence of each integrant (or non-integrated dsRNA molecule if using a non-library based method) is amplified by PCR using primers containing the sequence mapping to the top strand of the T7 promoter (or any other promoter used to express the dsRNA). Amplified DNA is then cloned into a cloning vector, such as pZERO blunt (Promega Corp.), and then sequenced. Sequences are compared to sequences in GenBank and/or other DNA databases to look for sequence identity or homology using standard computer programs. If the target mRNA remains unknown, the mRNA is cloned from the target cell line using primers derived from the cloned dsRNA by established techniques (Sambrook *et al.*, *supra*). Target validation is then carried out as described herein.

[0141] In the stably integrated dsRNA expression system described above, despite efforts to reduce negative position effects, inefficient dsRNA synthesis by PCR methods may occur. This can be circumvented by rescuing the integrated cDNA or randomized nucleic sequences into replicating plasmids. Rescued plasmids are amenable to amplification in bacteria and to sequencing. Rescue is achieved by re-transfecting the population of cells transfected with the dsRNA expression library with the rescue plasmid and a plasmid encoding Cre recombinase. The rescue plasmid carries a bacterial origin of replication, a bacterial antibiotic selection marker, an SV40 origin of replication, and an SV40 T antigen expression cassette, as well as *loxP* sites positioned as an inverted repeat to allow Cre-mediated double recombination. The SV40-based origin of replication in the rescue plasmid allows amplification of rescued sequences in the integrated cells. Following rescue, higher levels of transcription are anticipated, thereby favoring dsRNA formation. The cells are then screened for modulations in cell function, target nucleic acid expression, or target polypeptide biological activity changes as described herein.

Example 7: Prevention of an interferon response during gene silencing

[0142] As discussed above, during the above-described screening methods, induction of an interferon response is not desired, as this would lead to cell death, anti-proliferative responses, and possibly to prevention of gene silencing. One of the components of an interferon response is the induction of the interferon-induced protein kinase PKR. Suppression of the interferon response and/or the PKR response is desired in the target cells. The dsRNA delivery methods described herein

are performed such that an interferon response is not included. It is recognized, however, that certain conditions might present with an induction of the interferon response. To prevent such a response, a number of other strategies may be employed with any of the above described screening methods to identify a nucleic acid that modulates cell function, gene expression, or the polypeptide biological activity of a cell, as described herein.

[0143] To prevent an interferon response in a system involving stable integration of the nucleic acid containing the dsRNA expression cassette, the vectors used to generate either the *loxP* integrant or the vector that encodes the dsRNA expression cassette are designed to contain sequences that encode proteins that block the PKR response, such as the Vaccinia virus protein E3 (Romano *et al.*, Molecular and Cellular Biology 18: 7304-7316, 1998; Accession No. M36339), or a cellular protein P58^{IPK}, which the influenza virus mobilizes to block PKR (Gale *et al.*, Microbiology and Molecular Biology Reviews 64:239-280, 2000; Accession No. XM_032882). Several other viral proteins have also been identified (*e.g.*, Hepatitis C E2; Accession No. S72725) and may be similarly used. These proteins can be expressed in the desired cell types or in animals through the use of any of a number of commercially available mammalian expression vectors or vertebrate expression vectors. Such vectors can be obtained from a number of different manufacturers including Invitrogen (Carlsbad, CA) Promega (Madison, WI), or Clontech (Palo Alto, CA). An example of such a vector is the pCI-neo Mammalian Expression Vector from Promega.

[0144] Regardless of whether nucleic acid encoding a dsRNA is stably integrated into a chromosome or is not integrated into a chromosome, the following methods may be used to prevent an interference response in any of the screening methods of the present invention. In one example of an interferon avoidance strategy, interferon and PKR responses are silenced in the target cells using a dsRNA species directed against the mRNAs that encode proteins involved in the response. Desirably interferon response promoters are silenced using dsRNA. Alternatively, the expression of proteins that bind the interferon response element is abolished using dsRNA techniques.

[0145] In an alternative strategy, interferon and PKR knockout cell lines are created through approaches utilizing expression cassettes that encode an antisense RNA and ribozymes directed to the cellular mRNAs that encode the proteins involved in the response. Knockout cells are created by standard gene knockout technologies using homologous recombination to alter target sequences, using homologous DNA alone, or as complexes of RecA protein and single stranded DNA homologous to the target sequence(s). Interferon response element (IRE) sequences, sequences that encode transcription factors that bind IRE sequences, the promoter and/or gene sequences that encode proteins in the PKR

and interferon response pathways are molecules that are targeted for knockout.

[0146] In yet another alternative, chimeric oligonucleotides may be used to alter target sequences. Methods for inhibiting expression of polypeptides through chimeric oligonucleotides are well known in the art (Igoucheva and Yoon, Human Gene Therapy 11:2307-2312, 2000).

Example 8: Functional screening for cell invasion

[0147] Cell invasion is one cell function that may be evaluated in the search for novel genes that are modulated using the methods described herein. Matrigel, a biological extracellular matrix, has properties similar to that of a reconstituted basement membrane and has been used to measure the invasive potential of tumor cells (Platet and Garcia, *supra*). Cells transfected with randomized or cDNA libraries that have been cloned into PTGS vectors are monitored for their capacity to invade matrigel invasion chambers. Cells that have taken up sequences unrelated to invasion invade the matrigel as efficiently as vector-transfected control cells. Cells experiencing PTGS of genes that are involved in cell invasion invade much less efficiently. If the dsRNA expression cassette is stably integrated in a chromosome, these cells are retrieved and second and third rounds of selection are carried out in order to isolate specific nucleic acid sequences relevant to cell invasion. The effect of these sequences on invasion is ultimately confirmed by their ability to block the formation of tumors in animal models.

[0148] Several human cell lines, for example, MDA-MB-231, used by Platet and Garcia (*supra*), SKBr3, and MCF-7ADR, a more metastatic variant of MCF-7. MDA-MB-231 breast cancer cells (obtained from the American Type Culture Collection) are also transfected with cDNA libraries or randomized nucleic acid libraries constructed into the vectors described above. Desirably all cells in this assay contain a single copy of a transfected gene, as described above.

[0149] Cells cultured in commercially available 24- or 96-well formatted systems are used to carry out the cell invasion assay. As this screening protocol relies upon repeated rounds of selection, it may be desirable to keep the cell numbers in each well low enough that enrichment is seen in each succeeding round, yet high enough to recover sufficient cells to culture within a reasonable time period. Therefore, culture conditions that result in invasion by greater than 50% of the cells and that still permit recovery from the surface of the matrigel are made optimal. Non-invasive (NIH3T3 cells) or poorly invasive (MCF7) cell lines are analyzed in parallel as negative controls for invasion.

[0150] Initially, triplicate cultures of half-log order dilutions from 10^2 to 10^6 cells per well are plated. Cells are then recovered by "scrubbing" with a sterile cotton swab in fresh culture media and are seeded into 96-well plates. The number of invasive cells in the matrigel is

quantified using either an MTT-based assay (Sasaki and Passaniti, Biotechniques 24:1038-1043, 1998) or a fluorescent indicator (Gohla *et al.*, Clin. Exp. Metastasis 14:451-458, 1996).

[0151] Once the appropriate cell densities for the assay have been empirically determined, stable transfected cells are plated in the matrigel cell invasion chambers. Each experiment includes the following controls: a sample of untransfected cells as a reference culture; untransfected cells treated with anti-invasive chemotherapeutic agents, such as taxol or doxorubicin, as a positive control for inhibition of invasion; cells transfected with empty vectors to confirm that the vector alone had no effects on invasion; and cells transfected with genes that are known to block invasion in this assay, such as estrogen receptor- α or TIMP-2 (Kohn *et al.*, Cancer Research 55:1856-1862, 1995; and Woodhouse *et al.*, Cancer (Supplement) 80:1529-1536, 1997).

[0152] Cells that fail to invade the matrigel are removed from each well to the corresponding wells of a 96-well plate and cultured until macroscopic colonies are visible. It is important to collect cells at more than one time point after plating, since the time it takes for PTGS to be effective may vary, and it may be that different genes are active at different times after plating. Once the cells are transferred to 96-well plates, they are diluted out and taken through successive rounds of re-screening in the invasion assay in order to expand and isolate cell lines with altered invasive ability. As the population becomes more and more enriched for cells with a non-invasive phenotype, the reduction in invasive cells in the matrigel can be better quantified via MTT or fluorescence assays. Ultimately, a large panel of cloned double-stable cell lines is generated.

[0153] This assay can also be carried out with cells into which a dsRNA is not stably integrated into a chromosome. The assay is conducted essentially as described above except that multiple rounds of selection and re-screening are not necessary since the cell is transfected with only one dsRNA species. Thus, the target(s) of the PTGS event is readily identifiable using the cloning and sequencing techniques described above.

Example 9: Downregulation of HIV using HIV-derived dsRNA and inhibitors of the interferon response pathway

[0154] During the course of HIV infection, the viral genome is reverse transcribed into a DNA template that is integrated into the host chromosome of infected dividing cells. The integrated copy is now a blueprint from which more HIV particles are made. Several cell lines that contain integrated copies of a defective HIV genome, HIVgpt (strain HXB2) have been created. The HIVgpt genome contains a deletion of the HIV envelope gene; all other HIV proteins are encoded. The plasmid used to create these cell lines, HIVgpt, was obtained from the

AIDS Research and Reference Reagent Program Catalog. Stably integrated cell lines were made with human rhabdomyosarcoma (RD) cells. The lines were made by transfecting cells with the plasmid followed by selection of cells in mycophenolic acid. The HIVgpt genome encodes a mycophenolic acid (MPA) resistance gene in place of the envelope gene and thereby confers resistance to MPA. Cells resistant to MPA were clonally amplified. The media from the cultured clonally expanded cells was assayed for the presence of p24 (an HIV gag polypeptide that is secreted extracellularly). All cell lines were positive for p24, as assessed using a p24 ELISA assay kit (Coulter, Fullerton, CA). The cell lines also make non-infectious particles which can be rescued into infectious particles by co-expression of an HIV envelope protein.

[0155] The HIVgpt cell lines are used as a model system with which to downregulate HIV expression via PTGS. Plasmids encoding a 600 nt sense RNA, a 600 nt antisense RNA, or a 600 bp double stranded RNA (dsRNA), mapping to the same coordinates of the gag gene of HIV strain HXB2 are used to transfect cells (the map from which the coordinates are based is found at GenBank Accession number K03455, HIV(HXB2), complete genome, and the gag RNAs used in this study map to coordinates 901-1500). Expression of the RNAs is from T7 RNA polymerase promoter(s) located at the 5' end of the gag sense strand, at the 5' end of the antisense strand, or at converging positions at the 5' ends of both the sense and anti-sense strands, respectively. These encoded RNAs are not designed to be able to make protein (*i.e.*, they do not have a cap, a poly A tail, or the native initiation codon). Transcription of the RNAs is catalyzed by T7 RNA polymerase, provided from a second co-transfected T7 RNA polymerase expression plasmid. Control plasmids expressing a similar sized sense RNA, antisense RNA, and dsRNA derived from the gD gene of an HSV2 genome are included as experimental controls (the map from which the coordinates are based is found at GenBank Accession number K01408, HSVgD2 gene, and the HSVgD RNAs used in this study map to coordinates 313-872).

[0156] Cells used in these studies are transfected with an expression plasmid encoding a gene product known to interfere with the dsRNA induced interferon response or with the PKR response, as described above. The cells are transfected with lipofectamine (Gibco-BRL) as a transfecting agent according to the manufacturer's instructions.

[0157] Two days after transfection, the cells are harvested and seeded into six-well plates and cultured to approximately 80 to 90% confluence. Cells are co-transfected with the T7 RNA polymerase expression plasmid and one of the RNA expression plasmids, such that one well of cells receives the T7 RNA polymerase expression plasmid and the gag sense RNA expression plasmid; one well of cells receives the T7 RNA polymerase expression plasmid and the gag antisense RNA expres-

sion plasmid; one well of cells receives the T7 RNA polymerase expression plasmid and the gag dsRNA expression plasmid; one well of cells receives the T7 RNA polymerase expression plasmid and the HSVgD sense RNA expression plasmid; one well of cells receives the T7 RNA polymerase expression plasmid and the HSVgD antisense RNA expression plasmid; and one well of cells receives the T7 RNA polymerase expression plasmid and the HSVgD dsRNA expression plasmid. Transfection is again mediated through lipofectamine (Gibco-BRL). There also is a control group of cells receiving no RNA. The cells are monitored for p24 synthesis over the course of several weeks. The cells are assayed both by measuring p24 in the media of cells (using the p24 ELISA kit from Coulter, according to the manufacturer's instructions) and by immunostaining fixed cells for p24 using a rabbit polyclonal anti-p24 sera and anti-rabbit IgG that is FITC conjugated (Sigma). None of the gD RNAs specifically shut down p24 synthesis. The double stranded gag RNA significantly down regulates p24. The sense and antisense have only a modest effect on p24 synthesis and some of the effect is predicted to be through the ability of the sense and antisense gag RNAs to generate low levels of dsRNA species.

Example 10: Downregulation of PSA expression in human Rhabdomyosarcoma cells using intracellular expression of dsRNA

[0158] RD cells transiently expressing prostate specific antigen (PSA) were transfected with a T7 RNA polymerase expression vector and T7 RNA expression vectors expressing PSA dsRNA, PSA sense RNA, PSA antisense RNA, or control RNAs. The ability of the expressed RNAs to downregulate PSA expression was assessed, as described further below.

*Creation of a transient PSA*expression line*

[0159] The ability to downregulate expression of PSA following the expression of a PSA specific double-stranded RNA (dsRNA) was demonstrated in a human rhabdomyosarcoma cell line. Since available PSA cell lines are difficult to work with (*i.e.*, they are hard to transfect, and the cells tend to clump), a human cell line transiently expressing PSA was created. To create these cells, human rhabdomyosarcoma cells were transiently transfected with a PSA plasmid-based expression vector, under conditions that result in transfection of greater than 95% of the cells. Transfection was mediated with lipofectamine transfecting reagent (Gibco-BRL) according to the manufacturer's instructions. Expression of PSA was directed by the HCMV-IE promoter and the SV40 polyadenylation signal (Fig. 2). PSA expression was measured in the supernatant of transfected cells using a PSA ELISA kit (Oncogene Science Diagnostics, Cambridge, MA). No PSA was detected in the untransfected parental cells while PSA was abundantly ex-

47

EP 1 229 134 A2

48

pressed in cells receiving the PSA expression vector.

Downregulation of PSA expression

[0160] The PSA expressing cell line was used as a model system with which to demonstrate the downregulation of PSA protein levels by PTGS. In these studies, plasmids encoding an approximately 600 nt sense RNA, a 600 nt antisense RNA or a 600 nt dsRNA derived from a PSA cDNA were used to transfect the PSA expressing cell line (Fig. 2). Expression of the RNAs was from a T7 RNA polymerase promoter(s) located at the 5' end of the PSA sense strand, at the 5' end of the PSA antisense strand, or at converging positions at the 5' ends of both the sense and antisense strands respectively (Fig. 2). These encoded RNAs are not designed to be able to make protein (they do not have a cap, or a poly A tail). Transcription of the RNAs was catalyzed by T7 RNA polymerase, provided by a co-transfected T7 RNA expression plasmid. Control plasmids expressing similar sized sense RNA, antisense RNA, and dsRNA derived from the glycoprotein D (gD) gene of an Herpes simplex 2 (HSV-2) genome, as described above were included as experimental controls.

[0161] Cells used in these studies can optionally be transfected with an expression plasmid encoding a gene product known to interfere with the dsRNA induced interferon response or with the PKR response, as described above. The cells are transfected with lipofectamine (Gibco-BRL) as a transfecting agent according to the manufacturer's instructions.

[0162] Human rhabdomyosarcoma cells were seeded into six-well plates and cultured to approximately 80 to 90% confluence. The cells were co-transfected with (A) the PSA expression plasmid; (B) one of the T7 RNA expression plasmids; and (C) the T7 RNA polymerase expression plasmid, such that all PSA expressing cells were transfected with the T7 RNA polymerase expression plasmid and one of the following: the T7 sense PSA RNA expression construct, the T7 antisense PSA RNA expression construct, the T7 dsRNA PSA expression construct, the sense HSVgD RNA expression construct, the antisense HSVgD expression construct, or the dsRNA HSVgD expression construct. Cells received identical amounts of the PSA expression plasmid and the T7 RNA expression plasmid. The amounts of the T7 RNA expression plasmids were also constant amongst the transfected cells. Total DNA per transfection was held constant at 2.5 µg DNA per one well of a six-well plate. In those transfections where there was no T7 RNA expression plasmid, an inert plasmid DNA was used as filler DNA. Transfection was mediated by lipofectamine (Gibco-BRL) according to the manufacturer's instructions. There was also a control group of untransfected cells, as well as an untreated PSA control group of cells transfected with only the PSA expression plasmid in combination with the T7 RNA polymerase expression plasmid.

[0163] PSA-expressing cells that were not transfected with the T7 RNA expression plasmid, as well as cells transfected with the T7 HSV2-gD RNA expression plasmid all expressed PSA abundantly and at comparable levels. Cells transfected with the sense, antisense, and ds PSA RNA expression plasmids all exhibited varying degrees of inhibition of PSA expression. A 5%-10% reduction in expression was seen in cells expressing the PSA sense RNA, a 50% reduction was seen in cells expressing the PSA antisense RNA and greater than 95% reduction was seen in the cells expressing the PSA dsRNA (Fig. 3). The inhibition was seen within two days after transfection and continued up until the last time point taken (one month later) at which point PSA levels were beginning to decline in the untreated cells and the experiment was terminated. The untreated PSA controls as well as cells transfected with the T7 HSV2-gD control RNA expression plasmids all expressed PSA abundantly and at comparable levels (Fig. 3), indicating the specificity of dsRNA effectors to silence gene expression. During the one of month culture, cells were expanded into larger cultures at routine intervals.

[0164] Although the PSA specific dsRNA induced significant inhibition of PSA expression, antisense and sense PSA RNAs also induced some level of inhibition. Antisense PSA RNA has the potential to form dsRNA by annealing with PSA mRNA. Therefore the inhibition seen with antisense RNA may be explained by both an antisense mechanism and a dsRNA induced inhibition. A critical intracellular concentration of both antisense RNA and mRNA is required to generate dsRNA. Since much less dsRNA is made in the antisense RNA expressing cells relative to those cells designed to make dsRNA, a lesser inhibition of PSA in the antisense RNA expressing cells is expected if the threshold dsRNA levels required for efficient silencing have not been reached in those cells. We have also demonstrated that a small amount of antisense RNA can be found in cells transfected with our expression vectors (approximately 0.2% the amount of mRNA steady state levels). Antisense expression is presumably driven by cryptic promoters on the non-coding plasmid DNA strand. The observed sense RNA inhibition could therefore also involve a dsRNA molecule. RNA from transfected but untreated cells could also be analyzed to determine if the low level expression of antisense RNAs in these cells results in the production of detectable dsRNA species. Some low level expression of PSA occurred in cells expressing PSA dsRNA. It is likely that some percentage of cells did not take up the dsRNA expression cassette or that the threshold levels of dsRNA were not reached in some cells. No cellular toxicity was seen with any of the dsRNAs generated by the RNA expression vectors suggesting that cytoplasmic expression of dsRNA does not induce the interferon response. In contrast, cell death is induced when certain concentrations of *in vitro* produced dsRNA is delivered to cells via transfection with certain cationic lipids.

[0165] In summary, these results indicate that (i) PSA derived dsRNA is much more efficient than PSA antisense RNA in down-regulating PSA expression, (ii) the down-regulation is sequence specific; only the PSA derived dsRNA and not the control HSV-2 derived dsRNA induced down-regulation of PSA, and (iii) there is no toxicity associated with the cytoplasmic expression of long (600 bp) dsRNA molecules. Additionally, these experiments are the first demonstration of dsRNA mediated down-regulation of gene expression in a human cell line.

Example 11: Intracellular expression of dsRNA does not induce the type 1 interferon response (RNA stress response).

[0166] Human rhabdomyosarcoma (RD) cells were transfected with various dsRNA expression vectors such that dsRNA was transcribed in the transfected cells as described in Example 10. Transcription of dsRNA occurred in the cells within 24 hours after transfection and continued for the duration of the thirty day experiment. Cells and the supernatants from transfected cells were analyzed during the course of the experiment for any evidence of RNA stress response induction. No evidence of RNA stress response induction by intracellular expressed dsRNA was observed. RD cells have been shown by us to be responsive to type 1 interferon, both alpha and beta, and thus RD cells are capable of mounting an RNA stress response. In addition, positive controls were included in these experiments. A positive control for these experiments is a method of delivering dsRNA which induces the RNA stress response. All positive controls induced the RNA stress response. These experiments are described further below.

Assays performed to identify RNA stress response induction

[0167] The following assays were performed to measure the induction of an RNA stress response: TUNEL assay to detect apoptotic cells, ELISA assays to detect the induction of alpha, beta and gamma interferon, ribosomal RNA fragmentation analysis to detect activation of 2'5'OAS, measurement of phosphorylated eIF2a as an indicator of PKR (protein kinase RNA inducible) activation, proliferation assays to detect changes in cellular proliferation, and microscopic analysis of cells to identify cellular cytopathic effects. Apoptosis, interferon induction, 2'5' OAS activation, PKR activation, anti-proliferative responses, and cytopathic effects are all indicators for the RNA stress response pathway.

Transfection of cells

[0168] Approximately 7×10^5 RD cells were seeded into individual wells of six-well plates. Cells were transfected when they reached about 90% confluency. Cells were transfected with a T7 RNA polymerase expression

construct and a T7 dsRNA expression construct. The T7 dsRNA expression constructs encode converging T7 promoters located on either side of a 600 bp sequence (Fig. 2). Controls included cells transfected with the T7 RNA expression construct alone so that no dsRNA is made in these cells. Total DNA per transfection was held constant at 2.5 μ g DNA per one well of a six-well plate. When the T7 RNA polymerase and T7 dsRNA expression vectors were both used, 1.25 μ g of each DNA was used per transfection. In those transfections where there was no T7 dsRNA expression construct, inert filler DNA was used to bring the total DNA to 2.5 μ g per transfection. Transfection was mediated using Lipofectamine (Invitrogen) according to the manufacturer's directions. The positive control transfections included poly(I)(C) RNA and *in vitro* transcribed dsRNA of 600 bp that were both complexed with Lipofectamine and transfected into cells. The cells were transfected with *in vitro* transcribed ssRNA complexed with Lipofectamine. 2.5 μ g of each RNA was used per transfection. Other controls included untreated cells. Cells were kept in culture for one month by expanding into larger flasks as the cell numbers increased.

ELISA assays

[0169] Supernatants were removed from the transfected and untreated cells at time points of 1, 2, 7, 17, and 48 hours and every several days for up to one month after the 48 hour time point. Collected supernatants were stored at -80°C until they were analyzed for the presence of alpha, beta, and gamma interferon using commercially available ELISA kits. The Interferon-alpha ELISA kit was obtained from ENDOGEN (Rockford, IL), the Interferon-Beta ELISA kit was obtained from RD1 (Flanders, NJ), and the Interferon-gamma ELISA kit was obtained from R&D Systems (Minneapolis, MN). ELISAs were all performed according to the manufacturer's directions. Alpha, beta, and gamma interferon were not detected at increased levels in cells expressing intracellular dsRNA compared to the corresponding levels in untreated cells. However, considerable levels of beta interferon were found in cells transfected with poly (I)(C) or with *in vitro* transcribed dsRNA and ssRNA. Alpha and beta interferon induction are associated with induction of the RNA stress response.

TUNEL assay

[0170] Cells were stained for the presence of apoptotic nuclei using a commercially available kit, TdT FragEL, DNA Fragmentation Detection Kit, In Situ Apoptosis Assay from Oncogene (Boston, MA). Cells were stained according to the manufacturer's directions. Cells were stained at 2 hours, 7 hours, 17 hours, 2 days, 3 days, 4 days, and 5 days after transfection. There was no evidence of apoptosis induced by intracellular expressed dsRNA at any of the time points analyzed. How-

51

EP 1 229 134 A2

52

ever, the majority of cells transfected with poly (I)(C) or with the *in vitro* transcribed dsRNA were apoptotic by 17 hours after transfection. No evidence of apoptosis was observed in the untreated cells or in cells transfected with ssRNA. Apoptosis is an end result of the induction of the RNA stress response pathway.

2'5'OAS activation

[0171] The activation of 2'5'OAS was determined by performing ribosomal RNA fragmentation analysis. Briefly, following transfection, total RNA was extracted from cells using standard procedures. RNA was extracted at the following time points: 2 hours, 7 hours, 17 hours, 48 hours, 3 days, 4 days, and 5 days after transfection. 5-10 µg RNA was analyzed for each sample. RNA samples were first denatured in formaldehyde/formamide RNA sample buffer at 65°C for 10 minutes prior to being electrophoresed through 0.5X TBE agarose gels. Ribosomal RNA was visualized by staining with ethidium bromide followed by ultraviolet transillumination. Ribosomal RNA fragmentation was observed in cells transfected with poly (I)(C) and with the *in vitro* transcribed dsRNA. No fragmentation was observed in the untreated control cells, cells transfected with ssRNA, or in cells expressing intracellular dsRNA. These results indicate that 2'5'OAS was not activated by dsRNA when it was made intracellularly. 2'5'OAS activation is associated with induction /activation of the RNA stress response pathway.

PKR activation

[0172] The activation of PKR was determined by measuring the levels of eIF2alpha phosphorylation. Briefly, cells were lysed at various times after transfection (2 hours, 7 hours, 19 hours, 48 hours, 3 days, 4 days, and 5 days after transfection) and analyzed for levels of phosphorylated and non-phosphorylated eIF2 alpha. The protocol for lysing cells can be found in the following reference: Zhang F. *et al.*, J. Biol. Chem. 276 (27):24946-58, 2001. This analysis was performed as described for detecting PKR phosphorylation except that antibodies specific for phosphorylated and non-phosphorylated eIF2alpha were used. These antibodies are available from Cell Signaling Technology (Beverly, MA).

Cytopathic effect and antiproliferative responses

[0173] Cytopathic effect was assayed by analyzing cells microscopically using a light microscope. Cells were analyzed at daily intervals throughout the course of the experiment. Cytopathic effect is defined as any or all of the following: cells detaching from surface of well/flask, cells rounding up, an increased number of vacuoles in transfected cells with respect to the control untreated cells, or differences in morphology of cells with

respect to the untreated control cells. No cytopathic effect was seen in those cells expressing dsRNA intracellularly. Severe cytopathic effects were seen in cells transfected with Poly (I)(C) or with dsRNA made *in vitro*. Cytopathic effect is associated with the RNA stress response.

[0174] Antiproliferative responses were assayed by measuring the division rate of cells. The division rate is determined by counting cell numbers using standard procedures. Cells were counted every few days for the duration of the experiment. No antiproliferative responses were seen in cells expressing dsRNA intracellularly. Antiproliferative responses are associated with the RNA stress response.

Summary of results

[0175] The results of the above assays indicate that intracellular expression of dsRNA does not induce the RNA stress response. The cells that were used for these experiments were competent for RNA stress response induction as was demonstrated by the ability of cationic lipid complexed poly(I)(C) and *in vitro* transcribed RNA to induce/activate all tested components of this response. In addition, the cells were found to be responsive to exogenously added interferon. These results imply that the cells used for these experiments are not defective in their ability to mount an RNA stress response and therefore can be used as predictors for other cells, both in cell culture and *in vivo* in animal models. This method described here, which does not induce the interferon stress response, has been found to induce PTGS. This method therefore provides a method to induce PTGS without inducing an undesired RNA stress response.

[0176] Although these results were generated using a vector that utilizes a T7 transcription system and therefore expresses dsRNA in the cytoplasm, the vector system can be changed to other systems that express dsRNA intracellularly. Similar results are expected with these expression systems. These systems include, but are not limited to, systems that express dsRNA or hairpin RNAs in the nucleus, in the nucleus followed by transport of the RNAs to the cytoplasm, or in the cytoplasm using non-T7 RNA polymerase based expression systems.

Example 12: Optimization of the concentrations and relative ratios of *in vitro* or *in vivo* produced dsRNA and delivery agent

[0177] The optimal concentrations and ratios of dsRNA to a delivery agent such as a cationic lipid, cationic surfactant, or local anesthetic can be readily determined to achieve low toxicity and to efficiently induce gene silencing using *in vitro* or *in vivo* produced dsRNA.

53

EP 1 229 134 A2

54

Summary of factors effecting nucleic acid/cationic lipid interactions

[0178] Cationic lipid DNA interactions are electrostatic. Electrostatic interactions are highly influenced by the ionic components of the medium. The ability to form stable complexes is also dependent upon the intermolecular interactions between the lipid molecules. At low concentrations, certain inter-lipid interactions are preferred; at higher lipid concentrations, rapid condensates are formed due to higher order interactions. Although local interactions are similar in both of these instances (e.g., phosphoryl groups in the DNA and the charged cationic head group), the long range and inter-lipid interactions are substantially different. Similarly, structurally diverse variants can be obtained simply by changing the charge ratio of the complex by mixing varying amounts of cationic lipid with fixed concentrations of the nucleic acid or *vice versa*. This variation in the structure of the complexes is evidenced by altered physical properties of the complexes (e.g., differences in octanol partitioning, mobility on density gradients, charge density of the particle, particle size, and transfectability of cells in culture and *in vivo*) (Pachuk *et al.* DNA Vaccines - Challenges in Delivery, Current Opinion in Molecular Therapeutics, 2(2) 188-198, 2000 and Pachuk *et al.*, BBA, 1468, 20-30, (2000)). Furthermore, different lipids, local anesthetics, and surfactants differ in their interactions between themselves, and therefore novel complexes can be formed with differing biophysical properties by using different lipids singularly or in combination. For each cell type, the following titration can be carried out to determine the optimal ratio and concentrations that result in complexes that do not induce the stress response or interferon response. At several of these concentrations PTGS is predicted to be induced; however, PTGS is most readily observed under conditions that result in highly diminished cytotoxicity.

Complex formation

[0179] dsRNA is either produced by *in vitro* transcription using the T7 promoter and polymerase or another RNA polymerase, such as an *E. coli* RNA polymerase. dsRNA can also be produced in an organism or cell using endogenous polymerases.

[0180] Concentrations of dsRNA, such as PSA-specific dsRNA, are varied from 1 pg to 10 µg. In some instances, 150 ng of a plasmid that encodes a reporter of interest (PSA) to be silenced may be co-mixed at a concentration between 10 ng and 10 µg. The concentration of cationic lipid, cationic surfactant, local anesthetic, or any other transfection facilitating agent that interacts with the nucleic acid electrostatically are varied at each of the dsRNA concentrations to yield charge ratios of 0.1 to 1000 (positive / negative) (*i.e.*, the ratio of positive charge from lipids or other delivery agents to negative charge from DNA or RNA). The complexes are prepared

in water or in buffer (e.g., phosphate, HEPES, citrate, Tris-HCl, Tris-glycine, malate, etc. at pH values that range from 4.0 to 8.5), may contain salt (e.g., 1 - 250 mM), and may contain glycerol, sucrose, trehalose, xylose, or other sugars (e.g., mono-, di-, or polysaccharide). The mixture is allowed to sit at room temperature, desirably for 30 minutes, and may be stored indefinitely. The complexes are premixed in serum free media. The nucleic acid and the transfecting reagent may be mixed either through direct addition or through a slow mixing process, such as across a dialyzing membrane or through the use of a microporous particle or a device that brings the two solutions together at a slow rate and at low concentrations. In some instances, the two interacting components are mixed at low concentrations, and the final complex is concentrated using a diafiltration or any other concentrating device. Alternatively, if the complexes are formed at high concentrations of either or both of the interacting components, the complexes may be diluted to form an ideal transfection mixture.

Transfection protocol and analysis of dsRNA stress response

[0181] Complexes are added to cells that are ~60-80% confluent in serum free media. The complexes are incubated for various times (e.g., 10 minutes to 24 hours) with the cells at 37°C and diluted with serum containing media or washed and replated in serum free media. The cells are monitored for toxicity and analyzed at various times for signs of dsRNA response (e.g., TUNNEL assay to detect nicked DNA, phosphorylation of EIF2α, induction and activation of 2'5' OAS, or interferon-α and -β). Transfection conditions that result in less than 50%, 25%, 10%, or 1% cytotoxicity or that result in a less than 20, 10, 5, 2, or 1.5-fold induction of a stress response are analyzed to determine if PTGS was efficiently induced.

Determining induction of PTGS

[0182] PSA protein levels are determined in cell culture media using standard methods. The data is normalized to the number of live cells in culture to determine the concentrations required to induce PTGS.

Results

[0183] Using the above method, cationic lipid complexes of dsRNA induced toxicity at certain ranges. With lipofectamine as the cationic lipid, positive to negative charge ratios greater than 10 did not produce any detectable toxicity at any of the concentrations of dsRNA tested and induced a high level of PTGS, resulting in highly decreased levels of PSA in the culture medium. The RNA concentration ranges tested were 1 pg to 100 ng with a constant amount of lipofectamine (10 µL of a 2 mg/mL solution from GIBCO-BRL Life Technologies,

55

EP 1 229 134 A2

56

Bethesda, MD).

Example 13: Method to avoid dsRNA mediated activation of the RNA stress response pathway

[0184] One or more components of the RNA stress response pathway can be mutated or inactivated to avoid induction/activation of the component(s) by dsRNA that is delivered to the cell or animal for the purpose of inducing PTGS. These components, such as those illustrated in Fig. 4, can be knocked out singularly or in combination.

[0185] Various standard methods can be used to knockout components of the RNA stress response pathway, such as PKR, human beta interferon Accession No. M25460, and/or 2'5'OAS (Accession No. NM_003733). Alternatively or additionally, one or more interferon response element (IRE) sequences can be mutated or deleted using a knockout construct designed based on the IRE consensus sequence (Ghislain, *et al.*, J Interferon Cytokine Res. 2001 Jun 21(6):379-88.), and/or one or more transcription factors that bind IRE sequences, such as STAT1 (Accession number XM_010893), can be mutated or deleted. These methods include the use of antisense DNA/RNA, ribozymes, or targeted gene knockout technology mediated by homologous recombination. One skilled in the art is able to design the appropriate antisense sequences, ribozymes, and vectors for targeted knockouts. For example, targeted knockouts may be prepared by any of the following standard methods: Shibata *et al.*, Proc Natl Acad Sci U S A. 2001 Jul 17;98(15):8425-32. Review., Muylers *et al.*, Trends Biochem Sci. 2001 May;26(5):325-31., Paul *et al.*, Mutat Res. 2001 Jun 5;486(1):11-9., Shcherbakova *et al.*, Mutat Res. 2000 Feb 16;459(1):65-71., Lantsov. Ideal gene therapy: approaches and prospects Mol Biol (Mosk). 1994 May-Jun;28(3):485-95., in Gene Transfer and Expression - A Laboratory Manual editor: Michael Kriegler, Publisher- WH Freeman & Co, New York, NY, pages 56-60, 1990).

[0186] Knockout cells can be readily identified either through the use of an antibiotic resistance marker which when transferred to the chromosome confers resistance to the cell or through the use of dsRNA itself. In particular, dsRNA (*e.g.*, a high concentration of dsRNA) induces apoptosis in wild-type cells while mutant cells survive dsRNA treatment because they cannot mount a stress response. Yet another approach involves performing the dsRNA-induced PTGS experiment in the presence of large concentrations of IRE (dsDNA) oligo, which is expected to titrate activated STAT proteins. These oligos can be delivered intracellularly using transfecting agents or electroporation.

[0187] In another method of preventing the interferon response, cells (*e.g.*, RD cells) are transfected with a T7 RNA polymerase expression vector and a T7 dsRNA expression vector encoding dsRNA homologous to the human protein kinase PKR cDNA (accession number

M35663) or homologous to the coding sequence of any other component in the RNA stress response pathway. In one particular example, dsRNA corresponding to nucleotides 190-2000 is encoded by the T7dsRNA expression vector. The expression vectors are similar to those described in Example 10 and shown in Fig. 2, except that the dsRNA encoding sequence is derived from the human protein kinase PKR cDNA. Transfection in RD cells is performed as described in Example 10. Within 2-5 days post-transfection, the cells are functionally PKR negative.

Other Embodiments

[0188] From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

[0189] All publication, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

Claims

1. A method for identifying a nucleic acid that modulates the function of a vertebrate cell, said method comprising the steps of:
 - (a) transforming a population of vertebrate cells with a double stranded RNA expression library, wherein at least two cells of said population of cells are each transformed with a different nucleic acid from said double stranded RNA expression library, wherein said transformed nucleic acid is capable of forming double stranded RNA, and wherein said transformation and said formation of double stranded RNA are carried out under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
 - (b) selecting a vertebrate cell in which said nucleic acid is expressed in said cell; and
 - (c) assaying for a modulation in the function of said cell, wherein said modulation identifies a nucleic acid that modulates the function of said vertebrate cell.
2. A method for identifying a nucleic acid that modulates expression of a target nucleic acid in a vertebrate cell, said method comprising the steps of:

57

EP 1 229 134 A2

58

- (a) transforming a population of vertebrate cells with a double stranded RNA expression library, wherein at least two cells of said population of cells are each transformed with a different nucleic acid from said double stranded RNA expression library, wherein said transformed nucleic acid is capable of forming double stranded RNA, and wherein said transformation and said formation of double stranded RNA are carried out under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
- (b) selecting a vertebrate cell in which said nucleic acid is expressed in said cell; and
- (c) assaying for a modulation in the expression of a target nucleic acid in said cell, wherein said modulation identifies a nucleic acid that modulates expression of said target nucleic acid.
3. A method for identifying a nucleic acid that modulates the biological activity of a target polypeptide in a vertebrate cell, said method comprising the steps of:
- (a) transforming a population of vertebrate cells with a double stranded RNA expression library, wherein at least two cells of said population of cells are each transformed with a different nucleic acid from said double stranded RNA expression library, wherein said transformed nucleic acid is capable of forming double stranded RNA, and wherein said transformation and said formation of double stranded RNA are carried out under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
- (b) selecting for a vertebrate cell in which said nucleic acid is expressed in said cell; and
- (c) assaying for a modulation in the biological activity of a target polypeptide in said cell, wherein said modulation identifies a nucleic acid that modulates the biological activity of said target polypeptide.
4. The method of claim 1, wherein at most one nucleic acid is stably integrated into a chromosome of each cell.
5. The method of claim 1, said method further comprising:
- (d) identifying said nucleic acid by amplifying said nucleic acid and sequencing said amplified nucleic acid.
6. The method of claim 1, wherein said double stranded RNA expression library comprises cDNAs derived from said cells.
7. The method of claim 1, wherein said double stranded RNA expression library comprises randomized nucleic acids.
8. The method of claim 1, wherein said double stranded RNA expression library is a nuclear double stranded RNA expression library.
9. The method of claim 1, wherein said double stranded RNA expression library is a cytoplasmic double stranded RNA expression library.
10. The method of claim 1, wherein said cell is a mammalian cell.
11. The method of claim 10, wherein said cell is a human cell.
12. The method of claim 10, wherein said cell is selected from the group consisting of a cancer cell, a cell of the immune system, a stem cell, a neuronal cell, a muscle cell, and an adipocyte.
13. The method of claim 1, wherein said nucleic acid is contained in a vector.
14. The method of claim 13, wherein said vector comprises an RNA polymerase II promoter, an RNA polymerase I promoter, an RNA polymerase III promoter, or a mitochondrial promoter.
15. The method of claim 13, wherein the sense strand and the anti-sense strand of said double stranded RNA are transcribed from the same nucleic acid using two convergent promoters.
16. The method of claim 13, wherein said nucleic acid comprises an inverted repeat such that upon transcription said nucleic acid forms a double stranded RNA.
17. The method of claim 1, wherein said assaying comprises measuring an event selected from the group consisting of cell motility, apoptosis, cell growth, cell invasion, vascularization, cell cycle events, cell differentiation, cell dedifferentiation, neuronal cell regeneration, and the ability of a cell to support viral replication.
18. The method of claim 1, wherein said double stranded RNA is between 5 and 100 nucleotides in length, inclusive.
19. The method of claim 1, wherein said double strand-

59

EP 1 229 134 A2

60

- ed RNA is at least 100 nucleotides in length.
20. The method of claim 19, wherein said double stranded RNA is at least 250 nucleotides in length.
21. The method of claim 20, wherein said double stranded RNA is at least 500 nucleotides in length.
22. The method of claim 21, wherein said double stranded RNA is at least 1000 nucleotides in length.
23. A method for identifying a nucleic acid that modulates the function of a vertebrate cell, said method comprising the steps of:
- (a) transforming a population of vertebrate cells with a double stranded RNA under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
 - (b) selecting a vertebrate cell in which said double stranded RNA is expressed; and
 - (c) assaying for a modulation in the function of said cell, wherein said modulation identifies a nucleic acid that modulates the function of said cell.
24. A method for identifying a nucleic acid that modulates expression of a target nucleic acid in a vertebrate cell, said method comprising the steps of:
- (a) transforming a population of vertebrate cells with a double stranded RNA under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
 - (b) selecting a vertebrate cell in which said double stranded RNA is expressed; and
 - (c) assaying for a modulation in the expression of a target nucleic acid in said cell, wherein said modulation identifies a nucleic acid that modulates expression of said target nucleic acid.
25. A method for identifying a nucleic acid that modulates the biological activity of a target polypeptide in a vertebrate cell, said method comprising the steps of:
- (a) transforming a population of vertebrate cells with a double stranded RNA under conditions that inhibit or prevent an interferon response or a double stranded RNA stress response;
 - (b) selecting a vertebrate cell in which said double stranded RNA is expressed; and
 - (c) assaying for a modulation in the biological activity of a target polypeptide in said cell, wherein said modulation identifies a nucleic acid that modulates the biological activity of said target polypeptide.
26. The method of claim 23, said method further comprising:
- (d) identifying said nucleic acid by amplifying said nucleic acid and sequencing said amplified nucleic acid.
27. The method of claim 23, wherein said double stranded RNA is derived from said cells.
28. The method of claim 23, wherein said double stranded RNA is between 5 and 100 nucleotides in length, inclusive.
29. The method of claim 23, wherein said double stranded RNA is at least 100 nucleotides in length.
30. The method of claim 29, wherein said double stranded RNA is at least 250 nucleotides in length.
31. The method of claim 30, wherein said double stranded RNA is at least 500 nucleotides in length.
32. The method of claim 31, wherein said double stranded RNA is at least 1000 nucleotides in length.
33. The method of claim 23, wherein each cell is transformed with at most one double stranded RNA.
34. The method of claim 23, wherein said cell is a mammalian cell.
35. The method of claim 34, wherein said cell is a human cell.
36. The method of claim 23, wherein said population is transformed with at least 100 different double stranded RNAs.
37. The method of claim 1 or 23, said method further comprising:
- (d) identifying said nucleic acid that modulates the function of said cell.
38. The method of claim 2 or 24, said method further comprising:
- (d) identifying said nucleic acid that modulates expression of said target nucleic acid.
39. The method of claim 3 or 25, said method further

61

EP 1 229 134 A2

62

comprising:

- (d) identifying said nucleic acid that modulates
the biological activity of said target polypeptide.

5

10

15

20

25

30

35

40

45

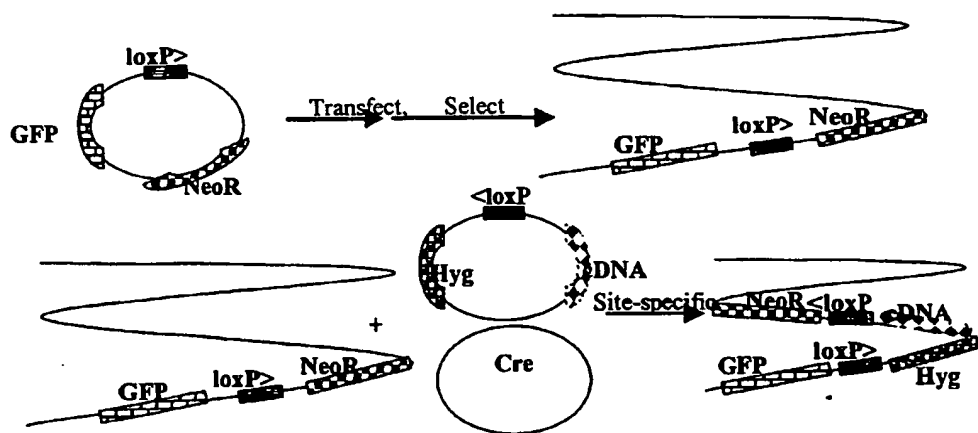
50

55

32

EP 1 229 134 A2

Fig. 1



EP 1 229 134 A2

Figure 2

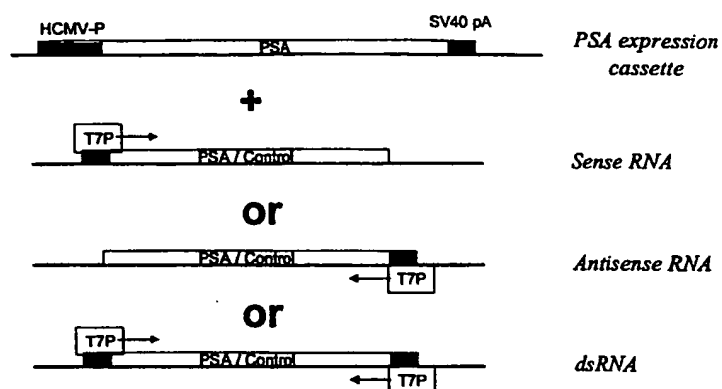
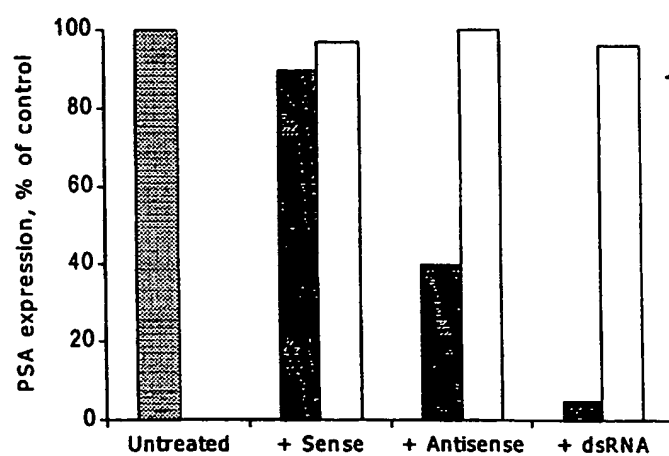


Figure 3



EP 1 229 134 A2

Figure 4

